

Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form is **not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at <https://eric.ed.gov/submit/> and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID

Publication/Completion Date—(if *In Press*, enter year accepted or completed)

Check type of content being submitted and complete one of the following in the box below:

- ☐ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

DOI or URL to published work (if available)

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

“This work was supported by U.S. Department of Education [Office name]
through [Grant number] to Institution] . The opinions expressed are
those of the authors and do not represent views of the [Office name]
or the U.S. Department of Education.



Published in final edited form as:

J Am Stat Assoc. 2017 ; 112(518): 794–805. doi:10.1080/01621459.2016.1173557.

Joint scale-change models for recurrent events and failure time

Gongjun Xu,

Assistant Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455

Sy Han Chiou,

Research Fellow, Department of Biostatistics, Harvard University, Boston, MA 02115

Chiung-Yu Huang,

Associate Professor, Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21205

Mei-Cheng Wang, and

Professor, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205

Jun Yan

Professor, Department of Statistics, University of Connecticut Storrs, CT 06269

Abstract

Recurrent event data arise frequently in various fields such as biomedical sciences, public health, engineering, and social sciences. In many instances, the observation of the recurrent event process can be stopped by the occurrence of a correlated failure event, such as treatment failure and death. In this article, we propose a joint scale-change model for the recurrent event process and the failure time, where a shared frailty variable is used to model the association between the two types of outcomes. In contrast to the popular Cox-type joint modeling approaches, the regression parameters in the proposed joint scale-change model have marginal interpretations. The proposed approach is robust in the sense that no parametric assumption is imposed on the distribution of the unobserved frailty and that we do not need the strong Poisson-type assumption for the recurrent event process. We establish consistency and asymptotic normality of the proposed semiparametric estimators under suitable regularity conditions. To estimate the corresponding variances of the estimators, we develop a computationally efficient resampling-based procedure. Simulation studies and an analysis of hospitalization data from the Danish Psychiatric Central Register illustrate the performance of the proposed method.

Keywords

Accelerated failure time model; Frailty; Informative censoring; Marginal models; Semiparametric methods

Supplement Materials

The supplement material contains additional simulation results and proofs of Lemmas 7.3, 7.4, and 7.5.

1 Introduction

Recurrent events arise frequently in various fields such as biomedical sciences, public health, engineering, and social sciences. Examples in medical research include acute rejection episodes after renal transplantation, cardiovascular events in survivors of myocardial infarction, and episodes of schizophrenia in chronic schizophrenic patients, which all contribute significantly to the overall disease burden and cost. Other important examples include car warranty claims, recidivism of sex offenders, machine failures, and many more (Cook and Lawless, 2007). A variety of statistical methods for recurrent event data have been proposed in the literature. The first class of models are the Cox-type models studied by Andersen and Gill (1982), Prentice et al. (1981), Pepe and Cai (1993), Lawless and Nadeau (1995), and Lin et al. (2000). These models postulate that the covariates effects are multiplicative on the intensity/rate function of the underlying recurrent event process, where the intensity function is the risk of experiencing a recurrent event conditional on the complete preceding event history and the rate function is the risk unconditional on the event history. Although the Cox-type models are very popular in regression analysis of recurrent event data, the proportionality assumption is often imposed by convention and may not be appropriate in some applications. Alternatively, Lin et al. (1998), Ghosh (2004), and Sun and Su (2008) extended the accelerated failure time (AFT) model and the accelerated hazard model for univariate survival time data to accommodate recurrent events by modeling the covariate effects on the time scale of the rate function and the cumulative rate function; Schaubel et al. (2006) studied additive rate models to allow for additive covariate effects. In general, the validity of the aforementioned methods relies heavily on the assumption that the censoring time distribution carries no information about the underlying recurrent event process.

In many instances, the observation of the recurrent event process can be stopped by the occurrence of a correlated failure event, where the failure event of interest can be a terminal event, such as death, or a non-terminal event, such as treatment failure due to drug resistance. Failing to account for the dependence between the recurrent events and the failure event may lead to substantial bias and misleading inferential results. In the literature, joint modeling approaches are commonly employed to analyze the recurrent events and the failure time data simultaneously. Among them, Lancaster and Intrator (1998) formulated a joint parametric model of repeated hospitalizations and survival time in HIV positive individuals. The intensity of repeated hospitalizations and the hazard of death were assumed to share a common baseline function and a gamma frailty. Conditioning on the frailty, the covariates have multiplicative effects on both the intensity function of recurrent hospitalizations and the hazard function of survival time. Other researchers, including Liu et al. (2004), Ye et al. (2007), and Liu and Huang (2009), considered semiparametric extensions of the joint model and developed Monte Carlo Expectation-Maximization algorithms to obtain the maximum likelihood estimators. Recently, Kalbfleisch et al. (2013) developed an iterative algorithm based on estimating equations to estimate the marginal and association parameters in a shared gamma frailty model without imposing a Poisson assumption on the recurrent event process. For non-Cox models, Zeng and Lin (2009)

considered a joint semiparametric transformation model and studied the semiparametric maximum likelihood estimation.

The shared-frailty models usually require a parametric assumption on the distribution of the frailty random variable so that the likelihood method can be used to estimate the joint model. Many authors have attempted to relax this assumption, as it can not be easily checked with the observed data. For Cox-type models, Wang et al. (2001), Huang and Wang (2004), and Huang et al. (2010) considered semiparametric estimation procedures that do not require the knowledge about the frailty; as a result, the dependence between recurrent events and failure event is left unspecified. While the estimation procedure studied by these authors are very flexible, a caveat is that the regression coefficients in the hazard function of the failure time usually do not have a simple marginal interpretation. In addition, their methods are based on the Poisson structure of the recurrent event process, which may not be satisfied in applications. On the other hand, the development of non-Cox models that allow for the association between recurrent and failure event to be nonparametric has been scarce. Zeng and Cai (2010) and Sun and Kang (2013) considered semiparametric models that postulate additive covariate effects on survivor's rate function (Luo et al., 2010). Parameters were estimated with estimating equation-based methods that allow for an arbitrary frailty distribution. Ghosh and Lin (2003) proposed a joint scale-change model for the marginal distributions of the recurrent event process and the failure time. At the cost of censoring some observed data artificially, their estimation procedure allows for an unspecified association between the two types of outcomes. Nonetheless, the estimation may be very inefficient because the artificial censoring can be heavy.

In this article, we introduce a joint scale-change model of the recurrent event process and the failure time. Assumeing that the covariate effects alter the time scale of the trajectories of recurrent event and failure time processes, the model specifies the AFT-type marginal models for the recurrent event and the failure time. As known in the literature, compared with the usual Cox-type models, a nice feature of the AFT-type model is that the regression parameters have a marginal interpretation (e.g., Ghosh and Lin, 2003). Although we employ a shared frailty to model the association between the subject-specific intensity of the recurrent events and the hazard of the failure event, the proposed approach is robust in the sense that no parametric assumption is imposed on the distribution of the unobserved frailty and that we do not require the strong Poisson-type assumption for the recurrent event process. Our estimation procedure does not require knowledge about the subject-specific frailty. The consistency and asymptotic normality of the estimators are established under suitable regularity conditions. We develop an efficient resampling-based sandwich estimator for the asymptotic variance of the estimator.

The rest of the article is organized as follows. The joint scale-change model is described in Section 2. Estimation procedure and variance estimator are proposed in Section 3, with proofs of large-sample properties given in the Appendix. A large-scale simulation study on performance of the estimators is reported in Section 4. An analysis of hospitalization data from the Danish Psychiatric Central Register is presented to illustrate the proposed methods in Section 5. A discussion concludes in Section 6.

2 Model Setup

Let $N(t)$ represent the number of events that occur over the interval $[0, t]$ and let D be the failure time of interest, such as the time of death or treatment failure. In most applications, the study subjects are followed for a limited period of time. Let C be the potential censoring time for reasons other than the failure event, and the observation of the failure time D is right censored by C . Suppose we are interested in making inference about the recurrent event process and the failure event in the time interval $[0, \tau]$, where the constant τ is determined with the knowledge that recurrent and failure events could potentially be observed up to time τ . Define $Y = \min(D, C)$ and $\Delta = I(D \leq C)$ so that the recurrent event process $N(\cdot)$ is observed up to the composite censoring time Y .

We introduce a nonnegative-valued latent frailty variable Z to account for the association between the underlying recurrent event process $N(\cdot)$ and the censoring events (D, C) . Let X be the a p -dimensional covariate vector for the recurrent event process and W be a q -dimensional covariate vector for the failure event D , where X and W may be completely distinct, overlapping, or even identical. We assume that, conditional on (Z, X, W) , $\{N(\cdot), D, C\}$ are mutually independent.

Conditioning on (Z, X, W) , we assume that the rate function of the counting process $N(\cdot)$, defined by $\lambda(t)dt = E[dN(t) | Z, X, W]$, takes the form

$$\lambda(t) = Z\lambda_0(te^{X^\top \alpha})e^{X^\top \alpha}, \quad t \in [0, \tau], \quad (1)$$

and the hazard function of the failure time D satisfies

$$h(t) = Zh_0(te^{W^\top \beta})e^{W^\top \beta}, \quad t \in [0, \tau], \quad (2)$$

where α and β are $p \times 1$ and $q \times 1$ vectors of parameters, respectively, and $\lambda_0(t)$ and $h_0(t)$ are absolutely continuous baseline intensity function and baseline hazard function, respectively. Denote the corresponding cumulative baseline intensity and baseline hazard functions by $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ and $H_0(t) = \int_0^t h_0(u) du$. For model identifiability, we assume that $\Lambda_0(\tau) = 1$ and $E(Z | X, W) = E(Z) = \mu_Z$ throughout the paper. Here the distribution function of the frailty random variable Z is left unspecified and the parameter μ_Z is to be estimated. Under the model assumptions (1) and (2), it is easy to see that the intensity of recurrent events, $\lambda(t)$, and the hazard of failure event, $h(t)$, can be both inflated (or deflated) by the shared subject-specific frailty Z .

As introduced, the failure event of interest can be a non-terminal event, such as treatment failure due to drug resistance, or a terminal event, such as death. For the former case, D can be viewed as an informative censoring time and Model (1) postulates the risk of recurrent event in the absence of any censoring and failure events. For the latter case, Model (1) can be interpreted in two ways depending on the research interest. First, as in Ghosh and Lin

(2003) and Huang and Wang (2004), recurrent events after the terminal event can be considered latent and modeled in (1) as if they could have occurred. Thus the cumulative rate function $\Lambda(t) = \int_0^t \lambda(u) du$, $t \in [0, \tau]$, gives the expected number of recurrent events that would occur in the time interval $[0, t]$ had the patient survived up to t . Alternatively, acknowledging that the terminal event precludes further occurrence of recurrent events after D , one can modify Model (1) as

$$\lambda(t) = Z\lambda_0(te^{X^\top\alpha})e^{X^\top\alpha}, \quad t \in [0, D]. \quad (3)$$

Under this minor modification, the cumulative baseline rate function $\Lambda_0(t)$, $t \in [0, \tau]$, does not possess a direct interpretation. Nevertheless, it still serves as an exploratory tool for evaluating covariate effects. We emphasize that the proposed estimation procedure in Section 3 does not require modeling recurrent events after the failure event and that it is valid under both Model (1) and Model (3).

The proposed model allows $\{N(\cdot), D\}$ to be correlated with C through the connection with unobserved frailty Z and covariates (X, W) . In other words, it allows the censoring time C to be correlated with the recurrent events and the failure event, and, hence, accommodates dependent censorship for both $N(\cdot)$ and D . The association among the triplet $\{N(\cdot), C, D\}$ is not completely specified because the distribution of Z is not specified in our model.

It is worth pointing out that, in contrast to the joint models considered by Huang and Wang (2004), Liu et al. (2004) and Ye et al. (2007), among others, the proposed model does not impose a Poisson-type assumption for the conditional distribution of $N(t)$ given Z . As will be shown in Sections 3 and 4, our proposed estimation procedure is robust against deviations from the Poisson assumption.

Another important feature of the proposed joint scale-change model is that it allows natural marginal interpretations for the regression parameters without having to specify a parametric distribution for the frailty variable. Under Model (1) the cumulative rate function of $N(\cdot)$ is given by

$$E\{N(t)|X, W\} = \mu_Z \Lambda_0(te^{X^\top\alpha}).$$

Thus the covariate effects modify the time scale of the cumulative rate function of the recurrent event process. If X is a treatment indicator, then the expected number of event by time t among treated subjects ($X = 1$) equals the expected number of events by time te^a in the control group ($X = 0$). Model (1) essentially specifies a AFT-type assumption for the recurrent event times. In particular, let t_{ij} be the time of the j th recurrent event for the i th subject and it can be shown that

$$\log t_{ij} = -X_i^\top \alpha + \varepsilon_{ij}, \quad (4)$$

where the error vectors $\varepsilon^j := (\varepsilon_{ij} : j = 1, \dots, i = 1, \dots, n)$, are independent and follow a common unspecified joint distribution (Ghosh and Lin, 2003). Note that, when the modified Model (3) is employed, the interpretation of (4) applies only to recurrent events that occur before the terminal event. A similar scale-change accelerated rate model as (1) without assuming the frailty variable has been proposed in Lin et al. (1998); however, the rank-based estimating equations considered there can not be applied to our setting due to the presence of informative censoring.

The marginal interpretation of Model (2) is clear if we consider a random variable D_0 which, given Z , has a hazard function $Zh_0(t)$. It can be shown that, unconditional on Z , the conditional distribution of D given (X, W) follows the AFT model $D = D_0 e^{-W^\top \beta}$, i.e.,

$$\log D = -W^\top \beta + \log D_0, \quad (5)$$

where the covariate effects alter the time scale of the failure event by a factor of $e^{-W^\top \beta}$. If W is a treatment indicator, then model (5) implies that the risk of experiencing the failure event at time t among the treated equals that at time $te^{-\beta}$ among the controls. In contrast, the regression parameters in the Cox-type joint models (e.g., Huang and Wang, 2004; Liu et al., 2004) do not have simple marginal interpretations. In particular, Huang and Wang (2004) assumed that, conditioning on the frailty, the conditional distribution of the recurrent event process follows a proportional intensity model and the failure time follows a proportional hazards model. As a result, the marginal distribution of $N(t)$ follows a proportional rate model, yet the marginal distribution of D usually does not follow the proportional hazards model.

3 Proposed Estimation Procedures

3.1 Estimation of α

We first present the estimation procedure for the regression parameter α in (1). Estimation of the regression parameter β in (2) will be considered in Section 3.2. Assume that the observed data $\{Y_i, \Delta_i, X_i, W_i, N_i(t), 0 \leq t \leq Y_i\}$, $i = 1, \dots, n$, on n subjects are independent and identically distributed realizations of $\{Y, \Delta, X, W, N(t), 0 \leq t \leq Y\}$. Let t_{ij} be the time of the j th recurrent event for the i th subject and let m_i be the number of recurrent events occurring before Y_i , that is, $m_i = \sum_{j=1}^{\infty} I(t_{ij} \leq Y_i)$. For a p -dimensional vector a , consider the transformation

$$t_{ij}^*(a) = t_{ij} e^{X_i^\top a} \text{ and } Y_i^*(a) = Y_i e^{X_i^\top a}.$$

Define the counting process for the transformed event times $N_i^*(t, a) = \sum_{j=1}^{\infty} I\{t_{ij}^*(a) \leq t\}$, so that $N_i^*(t, a) = N_i(te^{-X_i^\top a})$. Under model (1), we have

$$E\{N_i(t)|X_i, Z_i\} = \int_0^t Z_i \lambda_0(ue^{X_i^\top \alpha}) e^{X_i^\top \alpha} du = Z_i \Lambda_0(te^{X_i^\top \alpha}),$$

and

$$E\{N_i^*(t, a)|X_i, Z_i\} = Z_i \Lambda_0(te^{X_i^\top (\alpha - a)}).$$

Setting $a = \alpha$, it can be shown that, given (X_i, Z_i) , the mean function of the counting process $N_i^*(t, \alpha)$ does not depend on X_i . This property is essential for the construction of a robust estimation procedure for the recurrent event model.

Note that the transformed counting process $N_i^*(t, \alpha)$ is subject to informative censoring $Y_i^*(\alpha)$ because the two processes share the same frailty variable Z_i . As a result, conventional methods for recurrent event data can not be applied to make inference about $N_i^*(t, \alpha)$ even when the value of α is known. In this paper, we propose an estimation procedure that does not require the knowledge about the unobserved frailty variable.

We first consider a consistent estimator for $\Lambda_0(t)$. For any fixed a , define the stochastic processes

$$Q_n(t; a) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} I\{t_{ij}^*(a) \leq t\} \text{ and } R_n(t; a) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} I\{t_{ij}^*(a) \leq t \leq Y_i^*(a)\}.$$

Denote by $Q(t; a)$ and $R(t; a)$ the limiting functions of $Q_n(t; a)$ and $R_n(t; a)$. Under Model (1), we can show that

$$\begin{aligned} Q(u; a) &= E\left[\sum_{j=1}^{m_i} I\{t_{ij}^*(a) \leq u\}\right] = E\{N(ue^{-X_i^\top a} \wedge Y_i)\} \\ &= E\left[E\left\{Z_i \Lambda_0(ue^{X_i^\top (\alpha - a)} \wedge Y_i e^{X_i^\top \alpha}) \mid Z_i, Y_i, X_i\right\}\right] \\ &= E\left(E\left[\int_0^u Z_i I\{Y_i^*(a) \geq v\} d\Lambda_0\{ve^{X_i(\alpha - a)}\} \mid Z_i, Y_i, X_i\right]\right) \\ &= E\left[\int_0^u Z_i I\{Y_i^*(a) \geq v\} d\Lambda_0\{ve^{X_i(\alpha - a)}\}\right], \end{aligned} \quad (6)$$

where “ \wedge ” in the second equation denotes the minimum, and

$$\begin{aligned} R(u; a) &= E[\sum_{j=1}^{m_i} I\{t_{ij}^*(a) \leq u \leq Y_i^*(a)\}] \\ &= E(E[I\{Y_i^*(a) \geq u\} N(ue^{-X_i^\top a}) | Y_i, Z_i, X_i]) \\ &= E[I(Y_i^*(a) \geq u) Z_i \Lambda_0\{ue^{X_i(\alpha-a)}\}]. \end{aligned} \quad (7)$$

When a equals the true parameter α , we have $Q(u; \alpha) = \int_0^u E\{Z_i I(Y_i^*(\alpha) \geq v)\} d\Lambda_0(v)$ and $R(u; \alpha) = E\{Z_i I(Y_i^*(\alpha) \geq u)\} \Lambda_0(u)$. It follows from $\Lambda_0(\tau) = 1$ that

$$-\log \Lambda_0(t) = \int_t^\tau \frac{d\Lambda_0(u)}{\Lambda_0(u)} = \int_t^\tau \frac{dQ(u; \alpha)}{R(u; \alpha)}. \quad (8)$$

Thus $\Lambda_0(t)$ can be consistently estimated by

$$\exp \left\{ - \int_t^\tau \frac{dQ_n(u; \alpha)}{R_n(u; \alpha)} \right\}. \quad (9)$$

Interestingly, (9) is asymptotically equivalent to the truncation product-limit estimator

$$\hat{\Lambda}_n(t, \alpha) = \prod_{s_{(l)} > t} \left\{ 1 - \frac{d_{(l)}(\alpha)}{R_{(l)}(\alpha)} \right\},$$

where $\{s_{(l)}\}$ are the ordered and distinct values of the event times $\{t_{ij}^*(\alpha)\}$, $d_{(l)}(\alpha)$ is the number of events occurring at $s_{(l)}$, and $R_{(l)}(\alpha)$ is the total number of events with recurrent event times and failure event times satisfying $\{t_{ij}^*(\alpha) \leq s_{(l)} \leq Y_i^*(\alpha)\}$. As shown in the Appendix, $|\log \hat{\Lambda}_n(t; \alpha) - \int_t^\tau R_n^{-1}(u; \alpha) dQ_n(u; \alpha)| = o_p(n^{-1/2})$ uniformly in t .

It is worthwhile to point out that the proposed estimator $\hat{\Lambda}_n(t, \alpha)$ can be constructed directly by applying the conditional likelihood argument in Wang et al. (2001) and Huang and Wang (2004) to eliminate nuisance parameters from the likelihood of the transformed data. However, the validity of the argument requires a Poisson assumption on the recurrent event process $N(t)$ given Z . On the other hand, our derivation of the estimation procedure does not depend on the Poisson assumption of the recurrent event process and therefore relaxes the model assumptions imposed by existing literature.

The preceding discussion assumes that the value of α is known. We now construct an estimating equation for α . It follows from $E\{m_i | X_i, Y_i^*(\alpha), Z_i\} = Z_i \Lambda_0\{Y_i^*(\alpha)\}$ that

$$E[m_i \Lambda_0^{-1}\{Y_i^*(\alpha)\} | X_i, Y_i^*(\alpha)] = \mu_Z$$

and thus

$$E \left(X_i \left[m_i \Lambda_0^{-1} \{Y_i^*(\alpha)\} - \mu_Z \right] \right) = 0.$$

The above equations suggest that μ_Z can be estimated by

$$\hat{\mu}_Z = \frac{1}{n} \sum_{i=1}^n m_i \hat{\Lambda}_n^{-1} \{Y_i^*(\alpha), \alpha\},$$

and an estimating equation for α can be given by

$$S_n(a) := n^{-1} \sum_{i=1}^n X_i \left[m_i \hat{\Lambda}_n^{-1} \{Y_i^*(a)\} - \frac{1}{n} \sum_{j=1}^n m_j \hat{\Lambda}_n^{-1} \{Y_j^*(a)\} \right] = 0, \quad (10)$$

where we write $\hat{\Lambda}_n \{Y_i^*(a), a\}$ as $\hat{\Lambda}_n \{Y_i^*(a)\}$ when there is no ambiguity. We propose to estimate α by solving (10), that is, the proposed estimator $\hat{\alpha}_n$ for α satisfies $S_n(\hat{\alpha}_n) = 0$.

Given the estimator $\hat{\alpha}_n$, we propose to estimate the baseline cumulative rate function $\Lambda_0(t)$ by $\hat{\Lambda}_n(t, \hat{\alpha}_n)$, i.e.,

$$\hat{\Lambda}_n(t, \hat{\alpha}_n) = \prod_{s_{(l)} > t} \left\{ 1 - \frac{d_{(l)}(\hat{\alpha}_n)}{R_{(l)}(\hat{\alpha}_n)} \right\}.$$

It is easy to see that the estimation procedure does not involve the unobserved frailty variable Z_i . In the Appendix we derive an asymptotic representation of $\hat{\Lambda}_n(t, a)$ and use this representation to study the asymptotic properties of $\hat{\alpha}_n$ and $\hat{\Lambda}_n(t, \hat{\alpha}_n)$. In particular, we show that $\sqrt{n}(\hat{\alpha}_n - \alpha)$ converges weakly to a zero-mean multivariate normal distribution and $\sqrt{n}\{\hat{\Lambda}_n(t, \hat{\alpha}_n) - \Lambda_0(t)\}$, $t \in [0, \tau]$, converges weakly to a zero-mean Gaussian process. See Section 3.3 for more details.

Remark 3.1—More generally, it follows from $E\{m_i | X_i, Y_i^*(\alpha), Z_i\} = Z_i \Lambda_0\{Y_i^*(\alpha)\}$ that $E \left[w_i m_i \Lambda_0^{-1} \{Y_i^*(\alpha)\} | X_i, Y_i^*(\alpha) \right] = E[w_i] \mu_Z$, where w_i is a weight function depending on $\{X_i, Y_i^*(\alpha)\}$ and satisfying some regularity conditions. The equality motivates the estimator for μ_Z .

$$\hat{\mu}_Z = \left(\sum_{k=1}^n w_k \right)^{-1} \sum_{j=1}^n w_j m_j \hat{\Lambda}_n^{-1} \{Y_j^*(\alpha)\}.$$

We can show that $E(w_i X_i [m_i \hat{\Lambda}_n^{-1} \{Y_i^*(\alpha)\} - \mu_Z]) = 0$. Hence α can be estimated by solving the following estimating equation:

$$\frac{1}{n} \sum_{i=1}^n w_i X_i \left[m_i \hat{\Lambda}_n^{-1} \{Y_i^*(a)\} - \left(\sum_{k=1}^n w_k \right)^{-1} \sum_{j=1}^n w_j m_j \hat{\Lambda}_n^{-1} \{Y_j^*(a)\} \right] = 0.$$

This gives a class of estimating equations for different choices of w_i . In particular, setting $w_i = 1$ leads to the estimating equation (10). Moreover, setting $w_i = \Lambda_0 \{Y_i^*(a)\}$ and replace it with $\hat{\Lambda}_n \{Y_i^*(a)\}$, we have the following estimating equation

$$n^{-1} \sum_{i=1}^n X_i \left\{ m_i - \hat{\Lambda}_n \{Y_i^*(a)\} \left(\sum_{k=1}^n \hat{\Lambda}_n \{Y_k^*(a)\} \right)^{-1} \sum_{j=1}^n m_j \right\} = 0.$$

The construction of the optimal weight functions deserves further investigation, but is beyond the scope of this paper.

3.2 Estimation of β

We now focus on the estimation of the regression parameter β in (2). Because model (2) implies the AFT model (5), a naive approach is to estimate β by solving the rank-based estimating equation for the AFT model (e.g., Lin et al., 1998; Jin et al., 2003)

$$\frac{1}{n} \sum_{i=1}^n \Delta_i \left[W_i - \frac{\sum_{j=1}^n W_j I \{Y_j^*(b) \geq Y_i^*(b)\}}{\sum_{j=1}^n I \{Y_j^*(b) \geq Y_i^*(b)\}} \right] = 0. \quad (11)$$

However, the validity of the estimation procedure relies on the independent censoring assumption (given covariates W_i 's) which may not hold when the censoring time C_i is allowed to be correlated with the failure time D_i through the unobserved frailty Z_i . Our simulation study in Section 4 shows that the naive approach can be substantially biased.

To estimate β , note that, conditioning on (Z_i, X_i, W_i) , the stochastic process

$M_i(t) = \Delta_i I \{Y_i^*(\beta) \leq t\} - \int_0^t I \{Y_i^*(\beta) \geq u\} Z_i h_0(u) du$ is a martingale. Solving the martingale estimating equations $\sum_{i=1}^n \int_0^t dM_i(u) = 0$ for $t \in [0, \tau]$ and $\sum_{i=1}^n \int_0^\tau W_i dM_i(t) = 0$ yields an estimator for the baseline hazard function

$$\hat{h}_0(t) dt = \frac{d[\sum_{i=1}^n \Delta_i I \{Y_i^*(\beta) \leq t\}]}{\sum_{i=1}^n Z_i I \{Y_i^*(\beta) \geq t\}}. \quad (12)$$

This gives the estimating equation for β

$$U_n(b) := \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[W_i - \frac{\sum_{j=1}^n W_j Z_j I\{Y_j^*(b) \geq u\}}{\sum_{j=1}^n Z_j I\{Y_j^*(b) \geq u\}} \right] d[\Delta_i I\{Y_i^*(b) \leq u\}] = 0. \quad (13)$$

The estimating equation (13) cannot be evaluated because it depends on the subject-specific frailty, Z_i , which is not observable under our setting.

We consider an estimation procedure that does not require the knowledge about Z_i 's. We first note that, for fixed b , $U_n(b)$ can be reexpressed as

$$U_n(b) = \frac{1}{n} \sum_{i=1}^n W_i \Delta_i I\{Y_i^*(b) \leq \tau\} - \int_0^\tau \frac{\sum_{j=1}^n W_j Z_j I\{Y_j^*(b) \geq u\}}{\sum_{j=1}^n Z_j I\{Y_j^*(b) \geq u\}} d \left[\frac{1}{n} \sum_{i=1}^n \Delta_i I\{Y_i^*(b) \leq u\} \right].$$

Thus $U_n(b)$ defines a mapping from the four empirical processes,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n W_i \Delta_i I\{Y_i^*(b) \leq \tau\}, \quad & \frac{1}{n} \sum_{i=1}^n \Delta_i I\{Y_i^*(b) \leq u\}, \\ \frac{1}{n} \sum_{i=1}^n Z_i I\{Y_i^*(b) \geq u\}, \quad & \frac{1}{n} \sum_{i=1}^n W_i Z_i I\{Y_i^*(b) \geq u\}, \end{aligned} \quad (14)$$

to the real vector space. Replacing the four empirical processes in $U_n(b)$ with their respective limits yields the limiting function of $U_n(b)$, which is given by

$$\mathcal{U}(b) = E[W \Delta I\{Y^*(b) \leq \tau\}] - \int_0^\tau \frac{E[W Z I\{Y^*(b) \geq u\}]}{E[Z I\{Y^*(b) \geq u\}]} dE[\Delta I\{Y^*(b) \leq u\}].$$

The key in the proposed estimation procedure is to replace the two unobserved empirical processes in (13) by consistent estimators that converges to the same limiting functions.

We propose to replace Z_i in (13) by

$$\hat{Z}_i = \frac{m_i}{\hat{\Lambda}_n\{Y_i^*(\hat{\alpha}_n)\}},$$

where $\hat{\alpha}_n$ and $\hat{\Lambda}_n$ are the estimates constructed in the previous section. It can be shown that the stochastic processes $n^{-1} \sum_{i=1}^n \hat{Z}_i I\{Y_i^*(b) \geq u\}$ and $n^{-1} \sum_{i=1}^n W_i \hat{Z}_i I\{Y_i^*(b) \geq u\}$ converges uniformly to the functions $E[Z I\{Y^*(b) \geq u\}]$ and $E[W Z I\{Y^*(b) \geq u\}]$, respectively. Therefore, we can estimate β with estimating equation

$$\hat{U}_n(b) := \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[W_i - \frac{\sum_{j=1}^n W_j \hat{Z}_j I\{Y_j^*(b) \geq u\}}{\sum_{j=1}^n \hat{Z}_j I\{Y_j^*(b) \geq u\}} \right] d[\Delta_i I\{Y_i^*(b) \leq u\}] = 0.$$

Let $\hat{\beta}_n$ be the solution to $\hat{U}_n(\hat{\beta}_n) = 0$. The stochastic process \hat{U}_n converges uniformly to $\mathcal{U}(b)$, which ensures the consistency property of $\hat{\beta}_n$. In Section 3.3 we show that under certain regularity conditions, $\sqrt{n}(\hat{\beta}_n - \beta)$ converges weakly to a zero-mean multivariate normal distribution.

Finally, we consider the estimation of the baseline cumulative hazard function. If Z 's were observed, it follows from (12) that the baseline cumulative hazard function

$H_0(t) = \int_0^t h_0(u) du$ can be estimated by the Breslow-type estimator

$$\int_0^t \frac{d[\sum_{i=1}^n \Delta_i I\{Y_i^*(\beta) \leq u\}]}{\sum_{i=1}^n Z_i I\{Y_i^*(\beta) \geq u\}}. \quad (15)$$

It is easy to see that (15) also defines a mapping of two empirical processes and converges almost surely to $H_0(t)$ under the conditional independence assumption of C and D given (X, W, Z) . Since Z is unobservable, we propose to replace Z with \hat{Z} and β with $\hat{\beta}_n$ in (15) to obtain

$$\hat{H}_n(t, \hat{\beta}_n) = \int_0^t \frac{d[\sum_{i=1}^n \Delta_i I\{Y_i^*(\hat{\beta}_n) \leq u\}]}{\sum_{i=1}^n \hat{Z}_i I\{Y_i^*(\hat{\beta}_n) \geq u\}}. \quad (16)$$

The limit of $\hat{H}_n(t, \hat{\beta}_n)$ can be shown to be the functional of the limits of the two processes in (16), that is, $\hat{H}_n(t, \hat{\beta}_n)$ converges to $H_0(t)$ almost surely. Furthermore, in the next section we show that $\sqrt{n}\{\hat{H}_n(t, \hat{\beta}_n) - H_0(t)\}$, $t \in [0, \tau]$, converges weakly to a zero-mean Gaussian process.

Remark 3.2—Following the similar arguments, one can construct weighted estimating equations

$$\frac{1}{n} \sum_{i=1}^n \Delta_i \phi\{b, Y_i^*(b)\} \left[W_i - \frac{\sum_{j=1}^n W_j \hat{Z}_j I\{Y_j^*(b) \geq Y_i^*(b)\}}{\sum_{j=1}^n \hat{Z}_j I\{Y_j^*(b) \geq Y_i^*(b)\}} \right] = 0,$$

where ϕ is some specified weight function. It is worthwhile to point out that setting $\phi(b, t) = 1$ yields the log-rank-type estimating equation presented in this section, while setting

$\phi(b, t) = n^{-1} \sum_{j=1}^n Z_j I\{Y_j^*(b) \geq t\}$ gives a Gehan-type estimating equation.

3.3 Asymptotic results and variance estimation

To study the large sample properties of the proposed estimators, we impose the following regularity conditions:

- (C1) $P(\min\{Y^*(\alpha), Y^*(\beta)\} \geq \tau, Z > 0) > 0$.
- (C2) X and W are uniformly bounded and $E(Z^2) < \infty$.
- (C3) The conditional density of censoring time given (Z, X, W) is continuous and uniformly bounded.
- (C4) The rate function λ_0 and the hazard function h_0 have bounded second derivatives. In addition, λ_0 is bounded away from zero on $[0, \tau]$.
- (C5) The matrices J and $B(\beta)$ defined in (19) and (22) in the Appendix are non-singular.

Conditions C1–C5 are mild assumptions that are often assumed in survival models. Conditions C3 and C4 imply that $E[ZI(Y^*(\beta) \geq t)]$ is a continuous function for $t \in [0, \tau]$. Condition C4 assumes bounded second derivatives of baseline functions $\lambda_0(t)$ and $h_0(t)$, which is usually required for the AFT model to evaluate the asymptotic covariance matrix. With the proposed conditions, we have the following asymptotic results.

Theorem 3.1 — *Under conditions C1–C5, $\sqrt{n}(\hat{\alpha}_n - \alpha)$ and $\sqrt{n}(\hat{\beta}_n - \beta)$ converge weakly to multivariate normal distributions with mean zero and covariance matrices specified in the Appendix. Furthermore, for the estimated baseline rate function and baseline hazard function, we have $\sqrt{n}\{\hat{\Lambda}_n(t, \hat{\alpha}_n) - \Lambda_0(t)\}$ and $\sqrt{n}\{\hat{H}_n(t, \hat{\beta}_n) - H_0(t)\}$, $t \in [0, \tau]$, converges weakly to mean-zero Gaussian processes.*

The asymptotic covariance matrices of $\sqrt{n}(\hat{\alpha}_n - \alpha)$ and $\sqrt{n}(\hat{\beta}_n - \beta)$ depend on the nuisance density functions of both the survival time and the censoring time. In particular, as shown in the Appendix, equation (10) has the following asymptotic representation uniformly in a neighborhood of α

$$n^{1/2}S_n(a) = n^{1/2} \sum_{i=1}^n e_i(\alpha) + Jn^{1/2}(a - \alpha) + o_p(1 + n^{1/2}\|a - \alpha\|), \quad (17)$$

where the slope matrix J and $e_f(\alpha)$ are defined in (19) and (21), respectively. As stated in the Appendix, the asymptotic variance matrix of $\sqrt{n}(\hat{\alpha}_n - \alpha)$, denoted by $\Sigma(\alpha)$, has a sandwich form $J^{-1}V_\alpha(J^{-1})^\top$, where $V_\alpha = E[e(\alpha)e(\alpha)^\top]$. Since J involves the unknown baseline rate function, direct estimation of J requires numerical approximation techniques, such as nonparametric kernel density estimation, which can sometimes be complicated and inefficient. Jin et al. (2003) developed a resampling method to estimate the variance covariance matrix without density estimations. Let ξ_i , $i = 1, \dots, n$, be independent and identically distributed positive random variables with $E(\xi_i) = \text{Var}(\xi_i) = 1$. Define a perturbed version of the estimating function (10) as

$$\hat{S}_n^*(a) = \frac{1}{n} \sum_{i=1}^n \xi_i X_i \left(m_i [\hat{\Lambda}_n^* \{Y_i^*(a)\}]^{-1} - \frac{1}{n} \sum_{j=1}^n \xi_j m_j [\hat{\Lambda}_n^* \{Y_j^*(a)\}]^{-1} \right),$$

where

$$\hat{\Lambda}_n^*(t) = \prod_{s_{(l)} > t} \left[1 - \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} \xi_i I\{t_{ij}^*(a) = s_{(l)}\}}{\sum_{i=1}^n \sum_{j=1}^{m_i} \xi_i I\{t_{ij}^*(a) \leq s_{(l)} \leq y_i^*(a)\}} \right].$$

Given a realization of (ξ_1, \dots, ξ_n) , the solution to $\hat{S}_n^*(a) = 0$, \hat{a}_n^* , provides one draw of \hat{a}_n from its asymptotic distribution. Thus, the asymptotic variance-covariance matrix can be estimated by the sample variance matrix of a large sample of \hat{a}_n^* s. This resampling approach avoids the density estimation and is straightforward in principal but the implementation is computationally intensive because it requires to solve the perturbed estimating equations a large number of times.

In this paper, we adapt an efficient resampling approach that requires neither density estimation nor intensive computation. In the context of our estimator, the basic idea is to use the sandwich form of the asymptotic variance matrix, $\Sigma(a) = J^{-1} V_a (J^{-1})^\top$, and estimate the two components, J and V_a , separately.

We first focus on the estimation of V_a . Given a realization of (ξ_1, \dots, ξ_n) , consider the perturbed estimating function $n^{-1/2} \hat{S}_n^*(\hat{a}_n)$ evaluated at the value of the estimator \hat{a}_n from the observed data. By a similar argument as in Zeng and Lin (2008), $n^{-1/2} \hat{S}_n^*(\hat{a}_n)$ conditional on the observed data has the same asymptotic distribution as $n^{-1/2} S_n(a)$ evaluated at the true parameter a , with the limiting covariance matrix V_a . A consistent estimator of V_a , denoted as V_{a_n} , then can be obtained by the sample variance of the perturbed replicates of $n^{-1/2} \hat{S}_n^*(\hat{a}_n)$. Note that here we only need to evaluate the perturbed estimating equations at \hat{a}_n a large number of times and do not require solving the perturbed estimating equations repeatedly. Therefore, compared with the aforementioned resampling approach, this proposed resampling approach is computationally more efficient.

Estimation of the slope matrix, J , is challenging due to non-smoothness of the estimating function. We again adapt the resampling strategy. It follows from the decomposition (17) that for a p -dimensional standard normal random vector γ independent of the data

$$n^{1/2} S_n(a + n^{-1/2} \gamma) - n^{1/2} S_n(a) = J\gamma + o_p(1).$$

With $S_n(\hat{a}_n) = 0$, we have $n^{1/2} S_n(\hat{a}_n + n^{-1/2} \gamma) = J\gamma + o_p(1)$. For a large realization of γ , the j th row of J can then be approximated by regressing the j th component of $n^{1/2} S_n(\hat{a}_n +$

$n^{-1/2}\gamma)$ on γ . This gives estimator J_n . The target sandwich variance $\Sigma(\alpha)$ is then estimated by $J_n^{-1}V_{\alpha n}(J_n^{-1})^\top$.

As in the case of $\hat{\alpha}$, we appeal to the efficient resampling approach for the variance-covariance matrix of $\sqrt{n}(\hat{\beta}_n - \beta)$. Define the perturbed estimating function of $\hat{U}_n(b)$ as

$$\hat{U}_n^*(b) = \frac{1}{n} \sum_{i=1}^n \Delta_i \xi_i \left[W_i - \frac{\sum_{j=1}^n \xi_j W_j \hat{Z}_j^* I\{Y_j^*(b) \geq Y_i^*(b)\}}{\sum_{j=1}^n \xi_j \hat{Z}_j^* I\{Y_j^*(b) \geq Y_i^*(b)\}} \right] = 0,$$

where $\hat{Z}_i^* = \xi_i m_i / \hat{\Lambda}_n\{Y_i^*(\hat{\alpha}_n)\}$. Note that here \hat{Z}_i^* is also perturbed by using the estimator $\hat{\Lambda}_n\{Y_i^*(\hat{\alpha}_n)\}$. Using the definitions in Appendix, the variance-covariance matrix can be expressed as $\Sigma(\beta)$ as $B^{-1}V_\beta(B^{-1})^\top$, where $V_\beta = E[\zeta_1(\beta)\zeta_1(\beta)^\top]$. Evaluating $\hat{U}_n^*(b)$ at $\hat{\beta}_n$ with a realization of (ξ_1, \dots, ξ_n) provides one bootstrap replicate of $\hat{U}_n^*(\hat{\beta}_n)$. With a large number of replicates, V_β can be estimated sample variance of the bootstrap sample. The estimation of B is similar to that of J only with $S_n(\hat{\alpha}_n + n^{-1/2}\gamma)$ replaced by $U_n(\hat{\beta}_n + n^{-1/2}\gamma)$.

4 Simulation

Extensive numerical studies were carried out to assess the performance of the proposed estimators. The recurrent event process and the failure time were, respectively, generated from model (1) with the baseline intensity function $\lambda_0(t) = 1/10$ and model (2) with the baseline hazard function $h_0(t) = t/400$. The recurrent and survival outcomes can be potentially observed up to $\tau = 10$. For each subject, the covariates $X_1 = W_1$ and $X_2 = W_2$ were independently generated from the Bernoulli distribution with $\Pr(X_1 = 1) = 0.5$ and the Uniform(0, 1) distribution. The subject-specific latent variable Z was generated from either a Geometric distribution with mean 10 or a gamma distribution with mean 10 and variance 50, abbreviated by Geom(0.1) and Gamma(2, 5), respectively. The regression parameters (α, β) were set at $\alpha = (-1, 1)^\top$, $\beta = (-1, 1)^\top$ and $\alpha = (-1, 0)^\top$, $\beta = (-1, 0)^\top$.

To evaluate the robustness of the proposed method against deviations from the Poisson assumption, two types of counting process were used to generate the recurrent events. Specifically, conditioning on Z , the interarrival times of the recurrent events were independently generated from either an exponential distribution with mean $10e^{-X^\top \alpha/Z}$ (Scenario I) or a uniform(0, $20e^{-X^\top \alpha/Z}$) distribution (Scenario II). Thus, under Scenario I, the individual counting process is a homogeneous Poisson process with intensity function $\lambda(t) = Ze^{X^\top \alpha}/10$. The potential censoring time, C , was generated from an exponential random variable with mean 10 when $X_1 = 1$ or an exponential random variable with mean $300/Z^2$ when $X_1 = 0$. Therefore, subjects with $X_1 = 0$ are more likely to be censored. Sample size was set to be either $n = 200$ or $n = 500$. For each combination, 1000 replicates were simulated. Under these setups, the proportion of subjects experiencing the failure event before loss to follow-up ranges from 11.7% to 25.9%, the average length of follow-up ranges from 4.5 to 5.3, and the average number of observed recurrent events ranges from 1.8 to 3.3.

Tables 1 and 2 summarize the empirical bias, empirical standard error, average of the standard error estimates, and the empirical coverage probability of the proposed estimators. The proposed estimators are nearly unbiased and robust against deviations from the Poisson assumption, as indicated by the small empirical bias in the estimated regression coefficients. The average standard errors, computed from the proposed efficient resampling variance estimator with 500 bootstrap replicates, are reasonably close to the empirical standard errors, indicating that the proposed variance estimator performs well. In the simulations with $\alpha = (-1, 0)^T$, $\beta = (-1, 0)^T$, the estimated regression parameters have larger standard errors because the associated average number of recurrent events and average death rate are smaller. Moreover, compared to the scenarios where $Z \sim \text{Geom}(0.1)$, the estimated regression parameters under $Z \sim \text{Gamma}(2, 5)$ tend to have smaller standard errors because $\text{Geom}(0.1)$ has a greater variability than $\text{Gamma}(2, 5)$. For most cases, the empirical coverage probabilities of the 95% confidence intervals are reasonably close to the nominal level (95%). The empirical coverage probabilities for $\hat{\alpha}$ are closer to the anticipated level of 95% than that for $\hat{\beta}$, suggesting that the normality approximation for β may require a larger sample size than that for α . The empirical averages of the estimated μ_Z are also reasonably close to the true value. Figures 2 and 3 show the estimates and the empirical pointwise 95% confidence intervals of the baseline cumulative intensity function and baseline cumulative hazard function. The averages of $\hat{\Lambda}_0(t)$ and $\hat{H}_0(t)$ are almost indistinguishable from the truth for all cases considered.

For comparison, Tables 1 and 2 also report results from the method of Ghosh and Lin (2003), which requires an independent censoring assumption for the failure time data. The associated estimates are biased. The consequences of violating the independent censoring assumption can be explained as follows. In the simulated control group with $X = 0$, subjects with higher frailty values tend to have a shorter observed failure times. Thus, risk sets are more likely to consist of healthier subjects at later time points. As a result, treatment effect associated with failure time, β , is overestimated by the naive rank-based estimation procedure for the AFT model, which is used as the initial step in the estimation procedure of Ghosh and Lin (2003). The estimator of Ghosh and Lin (2003) is not able to recover α and the 95% coverage probabilities range from 31% to 98%. When the baseline intensity function is not constant in time, the performance of the proposed method is similar and the results are reported in the supplementary material.

5 The Danish Psychiatric Central Register data

The Danish Psychiatric Central Register (DPCR) (Munk-Jørgensen and Mortensen, 1997) records all admissions to psychiatric hospitals in the entire nation of Denmark, where participation is mandatory for all Danish psychiatric hospitals, relevant clinical departments, and units treating patients with schizophrenia. The register covers psychiatric inpatient services in Denmark since 1969 as well as outpatient contacts from 1994. The objective of DPCR is to document, monitor, and improve diagnosis and care provided by the Danish psychiatric health care system among patients with schizophrenia. Due to its national coverage, the DPCR has become a cost-effective tool for epidemiology studies to investigate prevalence and incidence of psychiatric disorders, morbidity and service use associated with

severe mental disorders, patterns of cares, and risk factors of mental disorder and treatment outcomes.

To illustrate the proposed methods, we analyzed admission records from 8811 individuals who had their first contact with Danish psychiatric services during the period between April 1, 1970 and March 25, 1988 with a diagnosis of schizophrenia. Of the 8811 patients, 5493 were male (62.3%) and 1065 (12.1%) had early-onset schizophrenia, defined as onset before 20 years of age. The time scale is time from study onset, i.e., the entry time of each patient into the study. After the initial contact, 82.4% (71.7%) of the early-onset (late-onset) schizophrenia patients experienced additional hospitalizations, while 74.7% (70.3%) of male (female) patients had additional hospitalizations. On average, each patient had 4.1 inpatient psychiatric admissions before leaving the study due to death or other reasons. A total of 1053 deaths, among whom 76 had early-onset schizophrenia and 368 were males, were observed before the end of follow-up.

The focus of our analysis is to evaluate the effects of onset age and gender on the rate of rehospitalization and the risk of death. We apply the proposed joint model to the DPCR data with onset age and gender as binary covariates: onset age (1 = early-onset, 0 = late-onset) and gender (1 = male, 0 = female). Table 3 summarizes the results of the data analysis with standard errors estimated from the efficient resampling method with 500 bootstrap replicates. Under Model (4), a positive (negative) regression coefficient implies a higher (lower) rate of rehospitalizations for $X = 1$ than those with $X = 0$. Similarly, under Model (5), a positive (negative) regression coefficient is associated with a shorter (longer) survival time for $X = 1$ as compared to those with $X = 0$. Controlling for the gender effect, early-onset is found to be associated with a higher risk of rehospitalization (p -value < 0.001) and a lower risk of death (p -value < 0.001). On average, the times to hospitalizations for a patient with early-onset schizophrenia are $\exp(-0.224) \approx 0.80$ times of those for a patient with late-onset schizophrenia. Similarly, the times to death for a early-onset patient are $\exp(0.911) \approx 2.49$ times of those for a late-onset patient. On the other hand, the times to hospitalizations for a male patient are 1.21 times of those for a female patient (lower risk of rehospitalization with p -value < 0.001), but the gender difference in the risk of death was not statistically significant (p -value 0.41). Similar findings were reported in Huang and Wang (2004) with a Cox-type joint model, but the interpretation of their estimators is based on the hazard function and, therefore, is not as direct as in the proposed model. Figure 1 shows the estimate of the baseline cumulative rate function and the baseline cumulative hazard function along with the corresponding pointwise 95% confidence intervals obtained by applying the proposed efficient resampling variance estimator with 500 bootstrap replicates. For comparison, we also evaluate the effects of onset age and gender on survival time under the estimator based on independent censoring assumption proposed by Ghosh and Lin (2003). As shown in Table 3, age and gender effects were not statistically significant for the rate of rehospitalization under the estimators of Ghosh and Lin (2003). As for the risk of death, the effect of early-onset is similar for the two models, but the direction of gender effects is opposite. The differences may be explained by the possible informative censoring C in this data.

6 Conclusion

In this paper, we propose a joint semiparametric model for the recurrent event process and the correlated failure time. The model assumes that the covariate effects alter the time scale of the trajectories of recurrent event and failure time processes. In particular, the joint model as in (1) and (2) specifies an AFT-type assumption for both the recurrent event and failure times. Therefore, as illustrated in (4) and (5), it allows a direct marginal interpretation for the regression parameters and provides an attractive alternative to the Cox-type models. A frailty variable is used to characterize the association between the subject-specific intensity of recurrent events and the hazard of failure event. It relaxes the independent censoring condition for observing both the recurrent event process and the failure time data, allowing informative censoring. In contrast to the conventional frailty models (Lancaster and Intrator, 1998; Liu et al., 2004; Ye et al., 2007; Liu and Huang, 2009), we do not impose parametric assumptions on the frailty distribution or the Poisson structure assumption on the recurrent event processes. To estimate the model parameters, we propose a borrow-strength procedure by first estimating the value of the latent variable from recurrent event data, then use the estimated value to construct estimating equations for the AFT model. A computationally efficient resampling method is proposed to estimate the variances of the model parameters. The proposed method does not require the estimation of neither the distribution of the frailty variable nor the baseline risk functions and therefore is easy to implement.

Some subtle issues are worth pondering depending on the primary interest of the research. Besides modeling the rate function approach as in (1), we can employ methods developed by other researchers to analyze recurrent event data with terminal event. For instance, Liu et al. (2004) and Ye et al. (2007) modeled the risk of recurrent event among survivors at time t ; Kurland and Heagerty (2005) considered a similar partly conditional model; Ghosh and Lin (2000) and Ghosh and Lin (2002) proposed to model the risk averaging over D .

Under our model specifications, the association between the recurrent and failure events is characterized by the unobserved frailty Z . The larger the variance of Z , the stronger the association. Compared with the approaches which parameterize the frailty distribution, the proposed method does not provide a direct measure on the degree of association between the recurrent event $N(t)$ and the failure event D . As an alternative, the size and shape association measures proposed by Wang and Huang (2014) may be applied to gain insight about the association between the two types of outcomes. Such association study will also be investigated.

For model identifiability, in this paper we assume $\Lambda_0(\tau) = 1$ while leaving μ_Z arbitrary. Alternatively we may impose $E(Z) = 1$ and allow $\Lambda_0(t)$ to be a bounded function on $[0, \tau]$. The proposed procedure is applicable for time-independent covariates. It would be also interesting to develop estimation procedures that allow for both time-invariant and time-dependent covariates. In addition, the proposed model assumes a common baseline intensity function for all subjects, which may not always be satisfied in practice. We may consider models that allow possible change after the occurrence of an event. Another interesting extension is to study the estimation equations with different weight functions as discussed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Huang and Wang's work was sponsored by National Institutes of Health grant R01CA193888.

References

- Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*. 1982; 10:1100–1120.
- Cook, RJ., Lawless, JF. *The statistical analysis of recurrent events*. Springer Science & Business Media; 2007.
- Ghosh D. Accelerated rates regression models for recurrent failure time data. *Lifetime Data Analysis*. 2004; 10:247–261. [PubMed: 15456106]
- Ghosh D, Lin D. Nonparametric analysis of recurrent events and death. *Biometrics*. 2000; 56:554–562. [PubMed: 10877316]
- Ghosh D, Lin D. Marginal regression models for recurrent and terminal events. *Statistica Sinica*. 2002; 12:663–688.
- Ghosh D, Lin DY. Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*. 2003; 59:877–885. [PubMed: 14969466]
- Huang C-Y, Qin J, Wang M-C. Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics*. 2010; 66:39–49. [PubMed: 19459839]
- Huang C-Y, Wang M-C. Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*. 2004; 99:1153–1165. [PubMed: 24068850]
- Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based Inference for the Accelerated Failure Time Model. *Biometrika*. 2003; 90:341–353.
- Jurečková J. Asymptotic linearity of a rank statistic in regression parameter. *The Annals of Mathematical Statistics*. 1969; 40:1889–1900.
- Jurečková J. Nonparametric estimate of regression coefficients. *The Annals of Mathematical Statistics*. 1971; 42:1328–1338.
- Kalbfleisch JD, Schaabel DE, Ye Y, Gong Q. An estimating function approach to the analysis of recurrent and terminal events. *Biometrics*. 2013; 69:366–374. [PubMed: 23651362]
- Kurland BF, Heagerty PJ. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics*. 2005; 6:241–258. [PubMed: 15772103]
- Lancaster T, Intrator O. Panel data with survival: hospitalization of HIV-positive patients. *Journal of the American Statistical Association*. 1998; 93:46–53.
- Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics*. 1995; 37:158–168.
- Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B*. 2000; 62:711–730.
- Lin DY, Wei LJ, Ying Z. Accelerated failure time models for counting processes. *Biometrika*. 1998; 85:605–618.
- Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C*. 2009; 58:65–81.
- Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics*. 2004; 60:747–756. [PubMed: 15339298]
- Luo X, Wang MC, Huang CY. A comparison of various rate functions of a recurrent event process in the presence of a terminal event. *Statistical Methods in Medical Research*. 2010; 19:167–182. [PubMed: 18445699]

- Munk-Jørgensen P, Mortensen PB. The Danish Psychiatric Central Register. Danish Medical Bulletin. 1997; 44:82–84. [PubMed: 9062767]
- Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. Journal of the American Statistical Association. 1993; 88:811–820.
- Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. Biometrika. 1981; 68:373–379.
- Schaubel DE, Zeng D, Cai J. A semiparametric additive rates model for recurrent event data. Lifetime Data Analysis. 2006; 12:389–406. [PubMed: 17031499]
- Sun L, Kang F. An additive-multiplicative rates model for recurrent event data with informative terminal event. Lifetime Data Analysis. 2013; 19:117–137. [PubMed: 22965680]
- Sun L, Su B. A class of accelerated means regression models for recurrent event data. Lifetime Data Analysis. 2008; 14:357–375. [PubMed: 18516715]
- Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. The Annals of Statistics. 1990; 18:354–372.
- van der Vaart, AW., Wellner, JA. Weak Convergence. Springer; 1996.
- Wang M-C, Huang C-Y. Statistical inference methods for recurrent event processes with shape and size parameters. Biometrika. 2014; 101:553–566. [PubMed: 26412863]
- Wang M-C, Qin J, Chiang C-T. Analyzing recurrent event data with informative censoring. Journal of the American Statistical Association. 2001; 96:1057–1065.
- Ye Y, Kalbfleisch JD, Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. Biometrics. 2007; 63:78–87. [PubMed: 17447932]
- Ying Z. A large sample study of rank estimation for censored regression data. The Annals of Statistics. 1993; 21:76–99.
- Zeng D, Cai J. A semiparametric additive rate model for recurrent events with an informative terminal event. Biometrika. 2010; 97:699–712. [PubMed: 23956458]
- Zeng D, Lin DY. Efficient resampling methods for nonsmooth estimating functions. Biostatistics. 2008; 9:355–363. [PubMed: 17925303]
- Zeng D, Lin D. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. Biometrics. 2009; 65:746–752. [PubMed: 18945267]

7 Appendix

We present the proofs for the asymptotic normality of

$\sqrt{n}\{\hat{\alpha}_n - \alpha\}$, $\sqrt{n}\{\hat{\Lambda}_n(t, \hat{\alpha}_n) - \Lambda_0(t)\}$, $\sqrt{n}\{\hat{\beta}_n - \beta\}$ and $\sqrt{n}\{\hat{H}_n(t, \hat{\beta}_n) - H_0(t)\}$. Proofs of some technical lemmas are given in the supplementary material due to space limitation.

Asymptotic linearity of $\hat{\Lambda}_n(t; a)$

We first show the asymptotic linearity of $\hat{\Lambda}_n(t; a)$, which will be used to establish the asymptotic normality of $\sqrt{n}(\hat{\alpha}_n - \alpha)$ and $\sqrt{n}\{\hat{\beta}_n - \beta\}$. Recall from (8) that

$$-\log \Lambda_0(t) = \int_t^\tau \Lambda_0^{-1}(u) d\Lambda_0(u) = \int_t^\tau R^{-1}(u; \alpha) dQ(u; \alpha).$$

Let $\Lambda_0(t, a) = \exp\{-\int_t^\tau R^{-1}(u; a) dQ(u; a)\}$. When $a = \alpha$, we have $\Lambda_0(t) = \Lambda_0(t, \alpha)$. By the uniform strong law of large numbers, we have $Q_n(u, a) \rightarrow Q(u, a)$ and $R_n(u, a) \rightarrow R(u, a)$ a.s. uniformly in a and $u \in [0, \tau]$. Furthermore we have $\|Q_n(u, a) - Q(u, a)\| = O_p(n^{-1/2})$ and $\|R_n(u, a) - R(u, a)\| = O_p(n^{-1/2})$, where $\|\cdot\|$ denotes the supremum norm. Then, we have the

uniform convergence result from the approximation results for the product-limit estimator:

$$\|-\log \hat{\Lambda}_n(t; a) - \int_t^\tau R_n^{-1}(u; a) dQ_n(u; a)\| = o_p(n^{-1/2}).$$

We first consider $\sqrt{n}\{\hat{\Lambda}_n(t; a) - \Lambda_0(t; a)\}$. We can show that

$$\begin{aligned} \int_t^\tau \frac{dQ_n(u; a)}{R_n(u; a)} &= \int_t^\tau \frac{dQ(u; a)}{R(u; a)} - \int_t^\tau \frac{\{R_n(u; a) - R(u; a)\}}{R(u; a)^2} dQ(u; a) + \int_t^\tau \frac{d\{Q_n(u; a) - Q(u; a)\}}{R(u; a)} + o_p(n^{-1/2}) \\ &= \int_t^\tau \frac{dQ(u; a)}{R(u; a)} - \frac{1}{n} \sum_{i=1}^n \eta_i(t; a) + o_p(n^{-1/2}), \end{aligned}$$

where

$$\eta_i(t; a) = \sum_{j=1}^{m_i} \int_t^\tau \frac{I\{t_{ij}^*(a) \leq u \leq Y_i^*(a)\} dQ(u; a)}{R(u; a)^2} - \sum_{j=1}^{m_i} \int_t^\tau \frac{dI\{t_{ij}^*(a) \leq u\}}{R(u; a)}.$$

Note that $E\eta_i = 0$. This gives $\log \Lambda_n(t, a) - \log \Lambda_0(t, a) = \frac{1}{n} \sum_{i=1}^n \eta_i(t; a) + o_p(n^{-1/2})$. Therefore, we have the asymptotic i.i.d. representation

$$\sqrt{n}\{\hat{\Lambda}_n(t; a) - \Lambda_0(t; a)\} = n^{-1/2} \sum_{i=1}^n \Lambda_0(t; a) \eta_i(t; a) + o_p(1). \quad (18)$$

This further implies the normality of $\sqrt{n}\{\hat{\Lambda}_n(t; a) - \Lambda_0(t; a)\}$.

We next show the asymptotic linearity results for $\sqrt{n}\{\hat{\Lambda}_n(t; a) - \hat{\Lambda}_n(t; \alpha)\}$. It can be verified that

$$\begin{aligned} &\sqrt{n} \left\{ \int_t^\tau \frac{dQ_n(u; a)}{R_n(u; a)} - \int_t^\tau \frac{dQ_n(u; \alpha)}{R_n(u; \alpha)} \right\} \\ &= \sqrt{n} \left\{ \int_t^\tau \frac{dQ(u; a)}{R(u; a)} - \int_t^\tau \frac{dQ(u; \alpha)}{R(u; \alpha)} \right\} \\ &\quad + \sqrt{n} \int_t^\tau \left\{ \frac{1}{R_n(u; a)} - \frac{1}{R(u; a)} - \frac{1}{R_n(u; \alpha)} + \frac{1}{R(u; \alpha)} \right\} dQ_n(u; a) \\ &\quad + \sqrt{n} \int_t^\tau \left\{ \frac{1}{R(u; a)} - \frac{1}{R(u; \alpha)} \right\} d\{Q_n(u; a) - Q(u; a)\} \\ &\quad + \sqrt{n} \int_t^\tau \left\{ \frac{1}{R_n(u; \alpha)} - \frac{1}{R(u; \alpha)} \right\} d\{Q(u; a) - Q(u; \alpha)\} \\ &\quad + \sqrt{n} \int_t^\tau \frac{1}{R_n(u; \alpha)} d\{Q_n(u; a) - Q(u; a) - Q_n(u; \alpha) + Q(u; \alpha)\} \\ &=: I_1 + I_2 + I_3 + I_4 + I_5. \end{aligned}$$

Applying a similar technique of Theorem 1 in Ying (1993), for some positive sequence $d_n \rightarrow 0$ and $\|a - \alpha\| \leq d_n$, we have uniformly in a and $u \in [0, \tau]$,

$$\frac{1}{R_n(u;a)} - \frac{1}{R(u;a)} - \frac{1}{R_n(u;\alpha)} + \frac{1}{R(u;\alpha)} = o_p(n^{-1/2}),$$

which implies that $I_2 = o_p(1)$ uniformly. In addition, by a similar argument as in Lemma 3 of Ying (1993), we have $I_3 + I_4 = o_p(1)$. Applying integration by parts and the following approximation that $Q_n(u; a) - Q(u; a) - Q_n(u; a) + Q(u; a) = o_p(n^{-1/2})$, we have $I_5 = o_p(1)$. Thus, $I_2 + I_3 + I_4 + I_5 = o_p(1)$ uniformly in $\|a - \alpha\| \leq d_n$ and $u \in [0, \tau]$.

Furthermore, from (6) and (7), we know

$$\begin{aligned} I_1 &= \int_t^\tau \frac{dQ(u;a)}{R(u;a)} - \int_t^\tau \frac{dQ(u;\alpha)}{R(u;\alpha)} \\ &= \int_t^\tau \frac{dQ(u;a) - E[ZI(Y^*(a) \geq u)] d\Lambda_0(u)}{R(u;a)} + \int_t^\tau \frac{E[ZI(Y^*(a) \geq u)] d\Lambda_0(u)}{R(u;a)} - \frac{E[ZI(Y^*(\alpha) \geq u)] d\Lambda_0(u)}{R(u;\alpha)} \\ &= \int_t^\tau \frac{E[X^\top ZI(Y^*(\alpha) \geq u)]}{R(u;\alpha)} d\{\lambda_0(u)u\} \cdot (\alpha - a) + \int_t^\tau \frac{\partial \{E[ZI(Y^*(a) \geq u)] R(u;a)^{-1}\}}{\partial a^\top} \Big|_{a=\alpha} d\Lambda_0(u) \cdot (\alpha - a) + o(\|\alpha - a\|), \end{aligned}$$

where the first part of the last equation follows from the approximation that

$$d\Lambda_0(u e^{X^\top(\alpha - a)}) - d\Lambda_0(u) = (1 + o(1)) X^\top(\alpha - a) d\{\lambda_0(u)u\}.$$

Thus, we have asymptotic approximation

$$\sqrt{n} \left\{ \int_t^\tau \frac{dQ(u;a)}{R(u;a)} - \int_t^\tau \frac{dQ(u;\alpha)}{R(u;\alpha)} \right\} = \kappa(t)^\top \sqrt{n}(a - \alpha) + o(n^{1/2}\|a - \alpha\| + 1),$$

where $\kappa(t)$ is the corresponding derivative matrix given by

$$\kappa(t) = \int_t^\tau \frac{E[XZI(Y^*(\alpha) \geq u)]}{R(u;a)} d\{\lambda_0(u)u\} + \int_t^\tau \frac{\partial E[ZI(Y^*(a) \geq u)] R(u;a)^{-1}}{\partial a} \Big|_{a=\alpha} d\Lambda_0(u).$$

Therefore, we have uniformly for $\|a - \alpha\| \leq d_n \rightarrow 0$ and $t \in [0, \tau]$

$$\sqrt{n} \left\{ \int_t^\tau \frac{dQ_n(u;a)}{R_n(u;a)} - \int_t^\tau \frac{dQ_n(u;\alpha)}{R_n(u;\alpha)} \right\} = \kappa(t)^\top \sqrt{n}(a - \alpha) + o_p(n^{1/2}\|a - \alpha\| + 1).$$

This implies $\sqrt{n}\{\hat{\Lambda}_n(t, a) - \hat{\Lambda}_n(t, \alpha)\} = \Lambda_0(t, \alpha) \kappa(t)^\top \sqrt{n}(a - \alpha) + o_p(n^{1/2}\|a - \alpha\| + 1)$, uniformly for $\|a - \alpha\| \leq d_n \rightarrow 0$ and $t \in [0, \tau]$.

Asymptotic results of $\hat{\alpha}_n$

We now prove the asymptotic normality of $\sqrt{n}(\hat{\alpha}_n - \alpha)$. Define $\mu_Z(a) = E[m\Lambda_0^{-1}\{Y^*(a)\}]$ and $s(a) = E[Xm\Lambda_0^{-1}\{Y^*(a)\}] - EX\mu_Z(a)$. Since $S_n(a) \rightarrow s(a)$ uniformly in a neighborhood of α and $s(a) \neq 0$ for $a \neq \alpha$, we have $\hat{\alpha}_n \rightarrow \alpha$. Next we show the asymptotic normality of $\hat{\alpha}_n$. For any sequence $d_n \rightarrow 0$, consider a in a neighborhood of α such that $\|a - \alpha\| \leq d_n$. We have

$$S_n(a) - S_n(\alpha) = \frac{1}{n} \sum_{i=1}^n X_i m_i \left[\frac{\hat{\Lambda}_n\{Y_i^*(\alpha)\} - \hat{\Lambda}_n\{Y_i^*(a)\}}{\hat{\Lambda}_n\{Y_i^*(a)\} \hat{\Lambda}_n\{Y_i^*(\alpha)\}} \right] - \frac{1}{n} \sum_{i=1}^n X_i [\hat{\mu}_Z(a) - \hat{\mu}_Z(\alpha)].$$

From the asymptotic linearity property of $\hat{\Lambda}_n$ developed in the last section, we have

$$\hat{\Lambda}_n\{Y_i^*(a)\} - \hat{\Lambda}_n\{Y_i^*(\alpha)\} = (1 + o_p(1)) \Lambda_0\{Y_i^*(\alpha)\} \kappa\{Y_i^*(\alpha)\}^\top Y_i e^{X_i^\top \alpha} X_i^\top (a - \alpha) + o_p(n^{-1/2}).$$

Further note that

$$\begin{aligned} \hat{\mu}_Z(a) - \hat{\mu}_Z(\alpha) &= \frac{1}{n} \sum_{i=1}^n m_i \left[\hat{\Lambda}_n^{-1}\{Y_i^*(a)\} - \hat{\Lambda}_n^{-1}\{Y_i^*(\alpha)\} \right] \\ &= \frac{1}{n} \sum_{i=1}^n m_i \left[\frac{\hat{\Lambda}_n\{Y_i^*(\alpha)\} - \hat{\Lambda}_n\{Y_i^*(a)\}}{\hat{\Lambda}_n\{Y_i^*(a)\} \hat{\Lambda}_n\{Y_i^*(\alpha)\}} \right]. \end{aligned}$$

Thus, uniformly in $\|a - \alpha\| \leq d_n$,

$$\begin{aligned} S_n(a) - S_n(\alpha) &= \frac{1}{n} \sum_{i=1}^n X_i m_i \Lambda_0^{-1}\{Y_i^*(\alpha)\} \kappa\{Y_i^*(\alpha)\}^\top Y_i e^{X_i^\top \alpha} X_i^\top (a - \alpha) + \frac{1}{n} \sum_{j=1}^n X_j \frac{1}{n} \sum_{i=1}^n m_i \Lambda_0^{-1}\{Y_i^*(\alpha)\} \kappa\{Y_i^*(\alpha)\}^\top Y_i e^{X_i^\top \alpha} X_i^\top (a - \alpha) + \\ &= J(a - \alpha) + o_p(n^{-1/2} + \|a - \alpha\|), \end{aligned}$$

where J is a $p \times p$ matrix defined as

$$J = E[Xm\Lambda_0^{-1}\{Y^*(\alpha)\} \kappa\{Y^*(\alpha)\}^\top Y^*(\alpha) X^\top] - E[X]E[m\Lambda_0^{-1}\{Y^*(\alpha)\} \kappa\{Y^*(\alpha)\}^\top Y^*(\alpha) X^\top].$$

(19)

This implies that $\sqrt{n}(\hat{\alpha}_n - \alpha) = J^{-1}n^{-1/2}S_n(\alpha) + o_p(1)$.

Next we show the normality of $n^{-1/2}S_n(a)$. $S_n(a)$ can be expressed as

$$S_n(\alpha) = \frac{1}{n} \sum_{i=1}^n X_i \left[m_i \hat{\Lambda}_n^{-1} \{Y_i^*(\alpha)\} - m_i \Lambda_0^{-1} \{Y_i^*(\alpha)\} \right] + \frac{1}{n} \sum_{i=1}^n X_i \left[m_i \Lambda_0^{-1} \{Y_i^*(\alpha)\} - \mu_Z \right] - \frac{1}{n} \sum_{i=1}^n X_i \{ \hat{\mu}_Z(\alpha) - \mu_Z \}.$$

(20)

From (18), the first term of the above display (20) equals

$$\begin{aligned} & -\frac{1}{n} \sum_{i=1}^n X_i m_i \frac{\hat{\Lambda}_n \{Y_i^*(\alpha)\} - \Lambda_0 \{Y_i^*(\alpha)\}}{\hat{\Lambda}_n \{Y_i^*(\alpha)\} \Lambda_0 \{Y_i^*(\alpha)\}} = \\ & -\frac{1}{n} \sum_{i=1}^n X_i m_i \frac{n^{-1} \sum_{j=1}^n \eta_j \{Y_i^*(\alpha)\}}{\Lambda_0 \{Y_i^*(\alpha)\}} + o_p(n^{-1/2}) \\ & = \\ & -\frac{1}{n} \sum_{i=1}^n \int \frac{x m \eta_i \{y^*(a)\}}{\Lambda_0 \{y^*(a)\}} dV(x, y, m) + o_p(n^{-1/2}), \end{aligned}$$

where $V(x, y, m)$ denotes the joint distribution function of (X, Y, m) . Similarly, the third term $\hat{\mu}_Z(\alpha) - \mu_Z$ equals

$$\begin{aligned} \hat{\mu}_Z(\alpha) - \mu_Z &= \hat{\mu}_Z(\alpha) - \frac{1}{n} \sum_{i=1}^n m_i \Lambda_0^{-1} \{Y_i^*(\alpha)\} + \frac{1}{n} \sum_{i=1}^n m_i \Lambda_0^{-1} \{Y_i^*(\alpha)\} - \mu_Z \\ &= \frac{1}{n} \sum_{i=1}^n \left[-\int \frac{m \eta_i \{y^*(a)\}}{\Lambda_0 \{y^*(a)\}} dV(x, y, m) + m_i \Lambda_0^{-1} \{Y_i^*(\alpha)\} - \mu_Z(a) \right] + o_p(n^{-1/2}), \end{aligned}$$

Combining the above results, we have $S_n(\alpha) = n^{-1} \sum_{i=1}^n e_i(\alpha) + o_p(n^{-1/2})$, where

$$e_i(\alpha) = \int \frac{(EX - x) m \eta_i \{y^*(\alpha)\}}{\Lambda_0 \{y^*(\alpha)\}} dV(x, y, m) + (X_i - EX) \left\{ \frac{m_i}{\Lambda_0 \{Y_i^*(\alpha)\}} - \mu_Z \right\}. \quad (21)$$

This gives the normality of $n^{1/2} S_n(\alpha)$ and the following approximation

$$n^{1/2}(\hat{\alpha}_n - \alpha) = n^{-1/2} J^{-1} \sum_{i=1}^n e_i(\alpha) + o_p(1).$$

This implies that $n^{1/2}(\hat{\alpha}_n - \alpha)$ converges weakly to a multivariate normal distribution with mean zero and variance $J^{-1} E[e_i(\alpha) e_i(\alpha)^T] (J^{-1})^T$.

Asymptotic results of $\hat{\Lambda}_n(t, \hat{\alpha}_n)$

The consistency of $\hat{\Lambda}_n(t, \hat{\alpha}_n)$ follows from that of $\hat{\alpha}_n$. From the asymptotic linearity results for $\sqrt{n}\{\hat{\Lambda}_n(t, \hat{\alpha}_n) - \hat{\Lambda}_n(t, \alpha)\}$ and $\sqrt{n}\{\hat{\Lambda}_n(t, \alpha) - \Lambda_0(t)\}$ in the preceding section, we have uniformly for $t \in [0, \tau]$

$$\begin{aligned}\sqrt{n}\{\hat{\Lambda}_n(t, \hat{\alpha}_n) - \Lambda_0(t)\} &= \sqrt{n}\{\hat{\Lambda}_n(t, \hat{\alpha}_n) - \hat{\Lambda}_n(t, \alpha)\} + \sqrt{n}\{\hat{\Lambda}_n(t, \alpha) - \Lambda_0(t)\} \\ &= n^{-1/2} \Lambda_0(t, \alpha) \sum_{i=1}^n \{\kappa(t)^\top J^{-1} e_i(\alpha) + \eta_i(t; \alpha)\} + o_p(1).\end{aligned}$$

Applying the functional central limit theorem, we have the weak convergence of $\sqrt{n}\{\hat{\Lambda}_n(t, \hat{\alpha}_n) - \Lambda_0(t)\}$ to a mean-zero Gaussian process for $t \in [0, \tau]$.

Asymptotic results of $\hat{\beta}_n$

We first show the consistency of $\hat{\beta}_n$. The first two empirical processes in equation (14) of the main file are sums of uncorrelated stochastic processes; hence we have the uniformly

convergence of $\|n^{-1} \sum_{i=1}^n W_i \Delta_i I\{Y_i^*(b) \leq \tau\} - E[W \Delta I\{Y^*(b) \leq \tau\}]\| \rightarrow 0$ and

$\|n^{-1} \sum_{i=1}^n \Delta_i I\{Y_i^*(b) \leq u\} - E[\Delta I\{Y^*(b) \leq u\}]\| \rightarrow 0$, a.s., following from the Glivenko-Cantelli theorem (van der Vaart and Wellner, 1996). From the asymptotic linearity of $\hat{\Lambda}_n$ and the consistency result of $\hat{\alpha}_n$, we have

$$\left\| \frac{\sum_{i=1}^n W_i \hat{Z}_i I\{Y_i^*(b) \geq u\}}{\sum_{i=1}^n \hat{Z}_i I\{Y_i^*(b) \geq u\}} - \frac{E[W Z I\{Y^*(b) \geq u\}]}{E[Z I\{Y^*(b) \geq u\}]} \right\| \rightarrow 0$$

in probability. The above results together imply that $\|\hat{U}_n(b) - \mathcal{U}(b)\| \rightarrow 0$ in probability. Since β is the unique solution to $\mathcal{U}(b) = 0$, we have the consistency of $\hat{\beta}_n$.

Next we show the normality of $\sqrt{n}(\hat{\beta}_n - \beta)$. We write

$$\mathcal{J}^{(k)}(u, \beta) = E \left[W^k Z I\{Y^*(\beta) \geq u\} \right]$$

for $k = 0, 1$ and 2 , where $W^0 = 1$, $W^1 = W$ and $W^2 = WW^\top$. Let

$$\tilde{N}_i(t) = \Delta_i I(Y_i \leq t) \text{ and } \tilde{W}(u, b) = \frac{\sum_{i=1}^n W_i \hat{Z}_i I\{Y_i^*(b) \geq u\}}{\sum_{i=1}^n \hat{Z}_i I\{Y_i^*(b) \geq u\}}.$$

Define the stochastic process $\tilde{M}_i(t) = \tilde{N}_i(t) - \int_0^t \hat{Z}_i h_0(u e^{W_i^\top \beta}) I\{Y_i \geq u\} du$. Note that

$\frac{1}{n} \sum_{i=1}^n \int_0^\tau \hat{Z}_i \{W_i - \tilde{W}(u, b)\} h_0(u) I\{Y_i^*(b) \geq u\} du = 0$. We can rewrite $\hat{U}_n(b)$ as

$$\hat{U}_n(b) = A_n(b) + B_n(b),$$

where

$$A_n(b) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{W_i - \tilde{W}(u, b)\} d\tilde{M}_i(ue^{-W_i^\top b}),$$

$$B_n(b) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \hat{Z}_i\{W_i - \tilde{W}(u, b)\} \{h_0(ue^{W_i^\top (\beta - b)}) - h_0(u)\} I\{Y_i^*(b) \geq u\} du.$$

To prove the normality, we introduce a local linear approximation of $\hat{U}_n(b)$. Let $\Phi(b)$ be defined by

$$\Phi(b) = \hat{U}_n(\beta) + B(\beta)(b - \beta),$$

where

$$B(\beta) = \int_0^\tau \left[\mathcal{S}^{(2)}(u, \beta) - \{\mathcal{S}^{(1)}(u, \beta)\}^2 / \mathcal{S}^{(0)}(u, \beta) \right] h'_0(u) du. \quad (22)$$

Let b^* be the solution to $\Phi(b) = 0$. A straightforward calculation shows that

$$\sqrt{n}(b^* - \beta) = -B^{-1}(\beta) \sqrt{n}\hat{U}_n(\beta).$$

Because $\sqrt{n}(\hat{\beta}_n - \beta) = \sqrt{n}(\hat{\beta}_n - b^*) + \sqrt{n}(b^* - \beta)$, the following two lemmas are sufficient to prove the normality of $\hat{\beta}_n$; see Jurečková (1969, 1971) and Tsiatis (1990) for details.

Lemma 7.1

As $n \rightarrow \infty$, $\sqrt{n}(b^* - \beta) \xrightarrow{d} N(0, \Sigma(\beta))$, where $\Sigma(\beta) = B^{-1}(\beta) E[\zeta_1(\beta)^2] \times B^{-1}(\beta)$ with ζ_1 defined as in the proof.

Lemma 7.2

For any constant $c > 0$, $\sup_{|b - \beta| \leq cn^{-1/2}} \sqrt{n}|\hat{U}_n(b) - \Phi(b)| \xrightarrow{p} 0$, as $n \rightarrow \infty$.

We proceed to prove the above lemmas.

Proof of Lemma 7.1—The weak convergence of

$\sqrt{n}(n^{-1} \sum_{i=1}^n W_i \Delta_i I\{Y_i^*(b) \leq \tau\} - E[W \Delta I\{Y^*(b) \leq \tau\}])$ and $\sqrt{n}(n^{-1} \sum_{i=1}^n \Delta_i I\{Y_i^*(b) \leq u\} - E[\Delta I\{Y^*(b) \leq u\}])$ follow from the classical central limit theorem and example 2.11.16 of van der Vaart and Wellner (1996).

Let $V(w, y, m)$ be the joint distribution function of (W, Y, m) . Using techniques similar to those in Huang and Wang (2004) and the approximation results for $\hat{\Lambda}_n(t, \hat{\alpha}_n)$ in the precede

section, we are able to establish the following asymptotic representations for the other two empirical processes

$$\begin{aligned} & \sqrt{n} \left[n^{-1} \sum_{i=1}^n W_i^k \hat{Z}_i I\{Y_i^*(\beta) \geq u\} - \mathcal{S}^{(k)}(u, \beta) \right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i^k I\{Y_i^*(\beta) \geq u\} m_i \left[\hat{\Lambda}_n^{-1}\{Y_i^*(\alpha)\} - \Lambda_0^{-1}\{Y_i^*(\alpha)\} \right] \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[W_i^k I\{Y_i^*(\beta) \geq u\} m_i \Lambda_0^{-1}\{Y_i^*(\alpha)\} - \mathcal{S}^{(k)}(u, \beta) \right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_i^{(k)}(u, \beta) + o_p(1), \end{aligned}$$

for $k = 0, 1$, where

$$\begin{aligned} \zeta_i^{(k)}(u, \beta) &= \int \frac{mw^k}{\Lambda_0(y^*(\alpha))} I\{y^*(\beta) \geq u\} \left[\kappa(y^*(\alpha))^\top J^{-1} e_i(\alpha) + \eta_i\{y^*(\alpha), \alpha\} \right] dV(w, y, m) \\ &+ \frac{m_i W_i^k}{\Lambda_0\{Y_i^*(\alpha)\}} I\{Y_i^*(\beta) \geq u\} \\ &- \mathcal{S}^{(k)}(u, \beta). \end{aligned}$$

Since $\mathcal{U}(\beta) = 0$ under the model assumption, we can derive the asymptotic representation

$$\begin{aligned} \sqrt{n} \hat{U}_n(\beta) &= n^{-1/2} \sum_{i=1}^n W_i \Delta_i I\{Y_i^*(\beta) \leq \tau\} - E[W \Delta I\{Y^*(\beta) \leq \tau\}] \\ &- \sqrt{n} \int_0^\tau \frac{\sum_{j=1}^n W_j \hat{Z}_j I\{Y_j^*(\beta) \geq u\}}{\sum_{k=1}^n \hat{Z}_k I\{Y_k^*(\beta) \geq u\}} d \left[\frac{1}{n} \sum_{i=1}^n \Delta_i I\{Y_i^*(\beta) \leq u\} \right] \\ &+ \sqrt{n} \int_0^\tau \frac{E[W Z I\{Y^*(\beta) \geq u\}]}{E[Z I\{Y^*(\beta) \geq u\}]} dE[\Delta I\{Y^*(\beta) \leq u\}] \\ &= n^{-1/2} \sum_{i=1}^n \zeta_i(\beta) + o_p(1), \end{aligned}$$

where

$$\begin{aligned} \zeta_i(\beta) &= W_i \Delta_i I\{Y_i^*(\beta) \leq \tau\} - E[W \Delta I\{Y^*(\beta) \leq \tau\}] \\ &+ \int_0^\tau \frac{\zeta_i^{(0)}(u, \beta) \mathcal{S}^{(1)}(u, \beta)}{\mathcal{S}^{(0)}(u, \beta)^2} dE[\Delta I\{Y^*(\beta) \leq u\}] - \int_0^\tau \frac{\zeta_i^{(1)}(u, \beta)}{\mathcal{S}^{(0)}(u, \beta)} dE[\Delta I\{Y^*(\beta) \leq u\}] \\ &- \int_0^\tau \frac{\mathcal{S}^{(1)}(u, \beta)}{\mathcal{S}^{(0)}(u, \beta)} d(\Delta_i I\{Y_i^*(\beta) \leq u\} - E[\Delta I\{Y^*(\beta) \leq u\}]). \end{aligned}$$

This gives the normality of $\sqrt{n} \hat{U}_n(\beta)$ which completes the proof.

Proof of Lemma 7.2—Note that $B_n(\beta) = 0$. For any sequence of real numbers b_n converging to β , $\sqrt{n}\{\hat{U}_n(b_n) - \Phi(b_n)\}$ can be decomposed into

$$\sqrt{n}\{\hat{U}_n(b_n) - \Phi(b_n)\} = \sqrt{n}\{A_n(b_n) - A_n(\beta)\} + \sqrt{n}\{B_n(b_n) - (b_n - \beta)B(\beta)\}. \quad (23)$$

We need the following lemmas whose proofs are in the supplementary material.

Lemma 7.3

For any sequence of real numbers b_n converging to β , $\sqrt{n}\{A_n(b_n) - A_n(\beta)\}$ converges to 0 in probability.

Lemma 7.4

Under the condition of Lemma 7.3,

$$B_n(b_n) = (b_n - \beta)\{B(\beta) + o_p(1)\}.$$

Therefore, we obtain that both terms in display (23) go to zero in probability by taking $b_n = \beta + dn^{-1/2}$. Thus, we have the pointwise uniform convergence of \hat{U} to Φ in the neighborhood of β , i.e., for any fixed $d > 0$,

$$\sqrt{n}\{\hat{U}_n(\beta + n^{-1/2}d) - \Phi(\beta + n^{-1/2}d)\} \xrightarrow{P} 0.$$

The tightness of the process $n^{1/2}\hat{U}_n(b)$ follows from the following lemma whose proof is postponed in the supplementary material

Lemma 7.5

For $|d| \leq c$ and any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{\beta + \frac{d}{\sqrt{n}} \leq b \leq \beta + \frac{d+\delta}{\sqrt{n}}} n^{1/2} |\hat{U}_n(b) - \hat{U}_n(\beta + dn^{-1/2})| \geq \varepsilon \right\} = 0.$$

This completes the proof.

Asymptotic results of $\hat{H}_n(t, \hat{\beta}_n)$

The consistency of $\hat{H}_n(t, \hat{\beta}_n)$ follows from that of $\hat{\beta}_n$. Furthermore, from a similar argument as in the proof of the normality of $\hat{\Lambda}_n(t, \hat{\alpha}_n)$, we have for b such that $\|b - \beta\| = O(n^{-1/2})$

$$\begin{aligned}
 \hat{H}_n(t, b) - \hat{H}_n(t, \beta) &= \int_0^t \frac{d[\sum_{i=1}^n \Delta_i I\{Y_i e^{W_i^\top b} \leq u\}]}{\sum_{i=1}^n \hat{Z}_i I\{Y_i e^{W_i^\top b} \geq u\}} - \int_0^t \frac{d[\sum_{i=1}^n \Delta_i I\{Y_i e^{W_i^\top \beta} \leq u\}]}{\sum_{i=1}^n \hat{Z}_i I\{Y_i e^{W_i^\top \beta} \geq u\}} \\
 &= \int_0^t \frac{dE[\Delta I\{Y e^{W^\top b} \leq u\}]}{\mathcal{S}^{(0)}(u, b)} - \int_0^t \frac{dE[\Delta I\{Y e^{W^\top \beta} \leq u\}]}{\mathcal{S}^{(0)}(u, \beta)} + o_p(n^{-1/2}) \\
 &= \int_0^t \frac{E[Z I\{Y e^{W^\top b} \geq u\} h_0(u e^{W^\top (\beta - b)}) e^{W^\top (\beta - b)}]}{\mathcal{S}^{(0)}(u, b)} du - \int_0^t h_0(u) du + o_p(n^{-1/2}) \\
 &= \int_0^t \frac{E[Z I\{Y e^{W^\top b} \geq u\} \{h_0(u e^{W^\top (\beta - b)}) e^{W^\top (\beta - b)} - h(u)\}]}{\mathcal{S}^{(0)}(u, b)} du + o_p(n^{-1/2}) \\
 &= - \int_0^t \frac{\mathcal{S}^{(1)}(u, \beta)^\top}{\mathcal{S}^{(0)}(u, \beta)} \{h'_0(u) u + h_0(u)\} du \cdot (b - \beta) + o_p(n^{-1/2})
 \end{aligned}$$

uniformly in b and $t \in [0, \tau]$, where $h'_0(t)$ is the derivative of $h_0(t)$. Therefore we have uniformly in $t \in [0, \tau]$,

$$\begin{aligned}
 n^{1/2} \{\hat{H}_n(t, \hat{\beta}) - H_0(t)\} &= n^{1/2} \{\hat{H}_n(t, \beta) - H_0(t)\} - \int_0^t \frac{\mathcal{S}^{(1)}(u, \beta)^\top}{\mathcal{S}^{(0)}(u, \beta)} \{h'_0(u) u + h_0(u)\} du \cdot n^{1/2}(\hat{\beta}_n - \beta) + o_p(n^{-1/2}).
 \end{aligned}
 \tag{24}$$

For the first term of the above display (24), we have that uniformly in $t \in [0, \tau]$

$$\begin{aligned}
 &n^{1/2} \{\hat{H}_n(t, \beta) - H_0(t)\} \\
 &= n^{-1/2} \int_0^t \frac{\sum_{i=1}^n [d\Delta_i I\{Y_i e^{W_i^\top \beta} \leq u\} - Z_i I\{Y_i e^{W_i^\top \beta} \geq u\} h_0(u) du]}{n^{-1} \sum_{i=1}^n Z_i I\{Y_i e^{W_i^\top \beta} \geq u\}} \\
 &\quad + n^{1/2} \int_0^t \left[\frac{\mathcal{S}^{(0)}(u, \beta) h_0(u)}{n^{-1} \sum_{i=1}^n \hat{Z}_i I\{Y_i e^{W_i^\top \beta} \geq u\}} - \frac{\mathcal{S}^{(0)}(u, \beta) h_0(u)}{\sum_{i=1}^n Z_i I\{Y_i e^{W_i^\top \beta} \geq u\}} \right] du \\
 &\quad + n^{1/2} \int_0^t \left[\frac{1}{n^{-1} \sum_{i=1}^n \hat{Z}_i I\{Y_i e^{W_i^\top \beta} \geq u\}} - \frac{1}{\sum_{i=1}^n Z_i I\{Y_i e^{W_i^\top \beta} \geq u\}} \right] \\
 &\quad \times n^{-1} \sum_{i=1}^n [d\Delta_i I\{Y_i e^{W_i^\top \beta} \leq u\} - \mathcal{S}^{(0)}(u, \beta) h_0(u) du] \\
 &= n^{-1/2} \sum_{i=1}^n \int_0^t \frac{1}{\mathcal{S}^{(0)}(u, \beta)} [d\Delta_i I\{Y_i e^{W_i^\top \beta} \leq u\} - Z_i I\{Y_i e^{W_i^\top \beta} \geq u\} h_0(u) du] \\
 &\quad + n^{-1/2} \sum_{i=1}^n \int_0^t \frac{\zeta_i^{(0)}(u, \beta)}{\mathcal{S}^{(0)}(u, \beta)} h_0(u) du + o_p(1),
 \end{aligned}$$

where in the last step we use the approximation result in the proof of Lemma 7.1 that

$$\sqrt{n} \left[n^{-1} \sum_{i=1}^n \hat{Z}_i I\{Y_i^*(\beta) \geq u\} - \mathcal{S}^{(0)}(u, \beta) \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_i^{(0)}(u, \beta) + o_p(1).$$

For the second term of (24), Lemma 7.1 implies that

$$n^{1/2}(\hat{\beta}_n - \beta) = -B^{-1}n^{-1/2} \sum_{i=1}^n \zeta_i(\beta) + o_p(1). \text{ Therefore we have}$$

$$n^{1/2}\{\hat{H}_n(t, \hat{\beta}) - H_0(t)\} = n^{-1/2} \sum_{i=1}^n \psi_i(t) + o_p(n^{-1/2}), \text{ with}$$

$$\begin{aligned} \psi_i(t) = & \int_0^t \frac{1}{\mathcal{J}^{(0)}(u, \beta)} \left[d\Delta_i I\{Y_i e^{W_i^\top \beta} \leq u\} - Z_i I\{Y_i e^{W_i^\top \beta} \geq u\} h_0(u) du \right] \\ & + \int_0^t \frac{\zeta_i^{(0)}(u, \beta)}{\mathcal{J}^{(0)}(u, \beta)} h_0(u) du + \int_0^t \frac{\mathcal{J}^{(1)}(u, \beta)^\top}{\mathcal{J}^{(0)}(u, \beta)} \{h_0'(u)u + h_0(u)\} du \cdot B^{-1} \zeta_i(\beta). \end{aligned}$$

By the functional central limit theorem, we have that $n^{1/2}\{\hat{H}_n(t, \hat{\beta}) - H_0(t)\}$ converges weakly to a mean-zero Gaussian process with covariance function $E[\psi_1(t)\psi_1(s)]$ for $t, s \in [0, \tau]$.

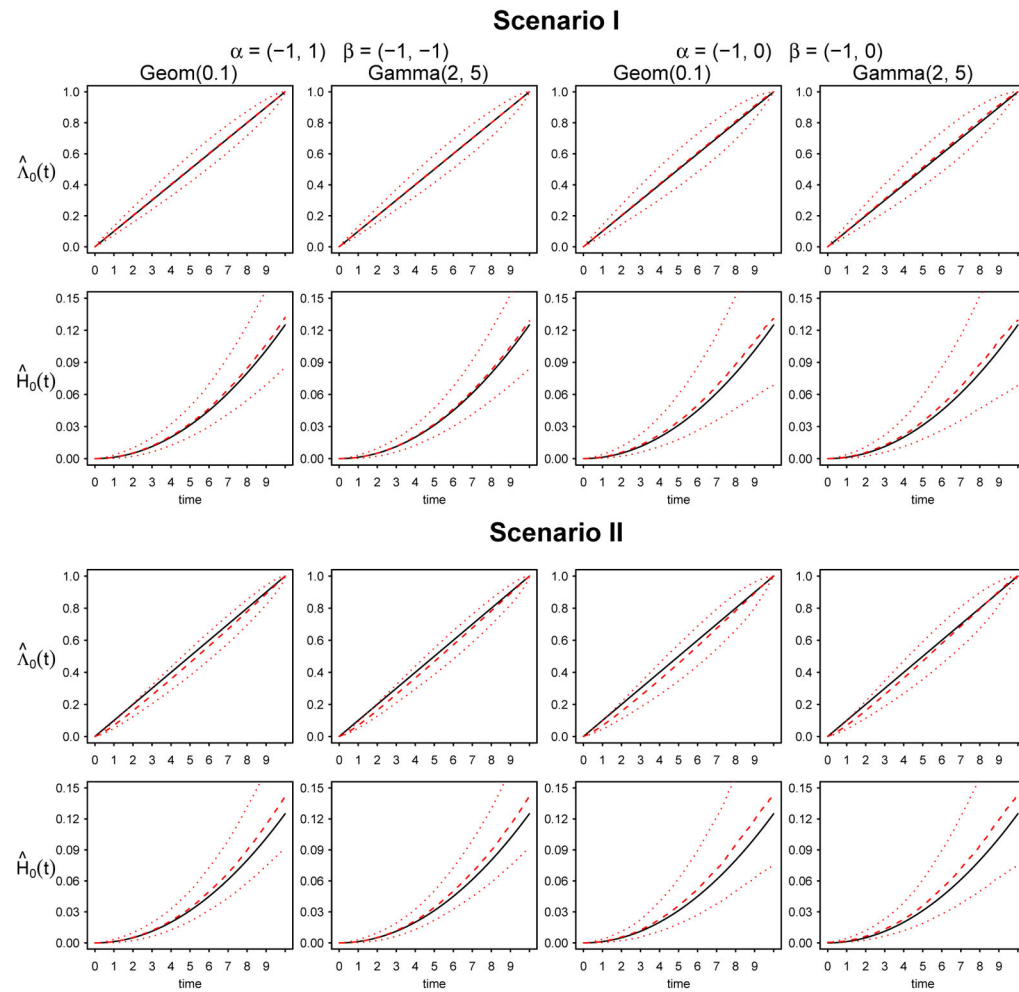


Figure 1.

Plots of estimated $\hat{\Lambda}_0(t)$ and $\hat{H}_0(t)$ with pointwise 95% confidence intervals for $n = 200$ (—, true curve; ---, empirical average; ----, pointwise 95% confidence intervals). Conditioning on Z , the recurrent process is a Poisson process and a non-Poisson process for Scenario I and Scenario II, respectively.

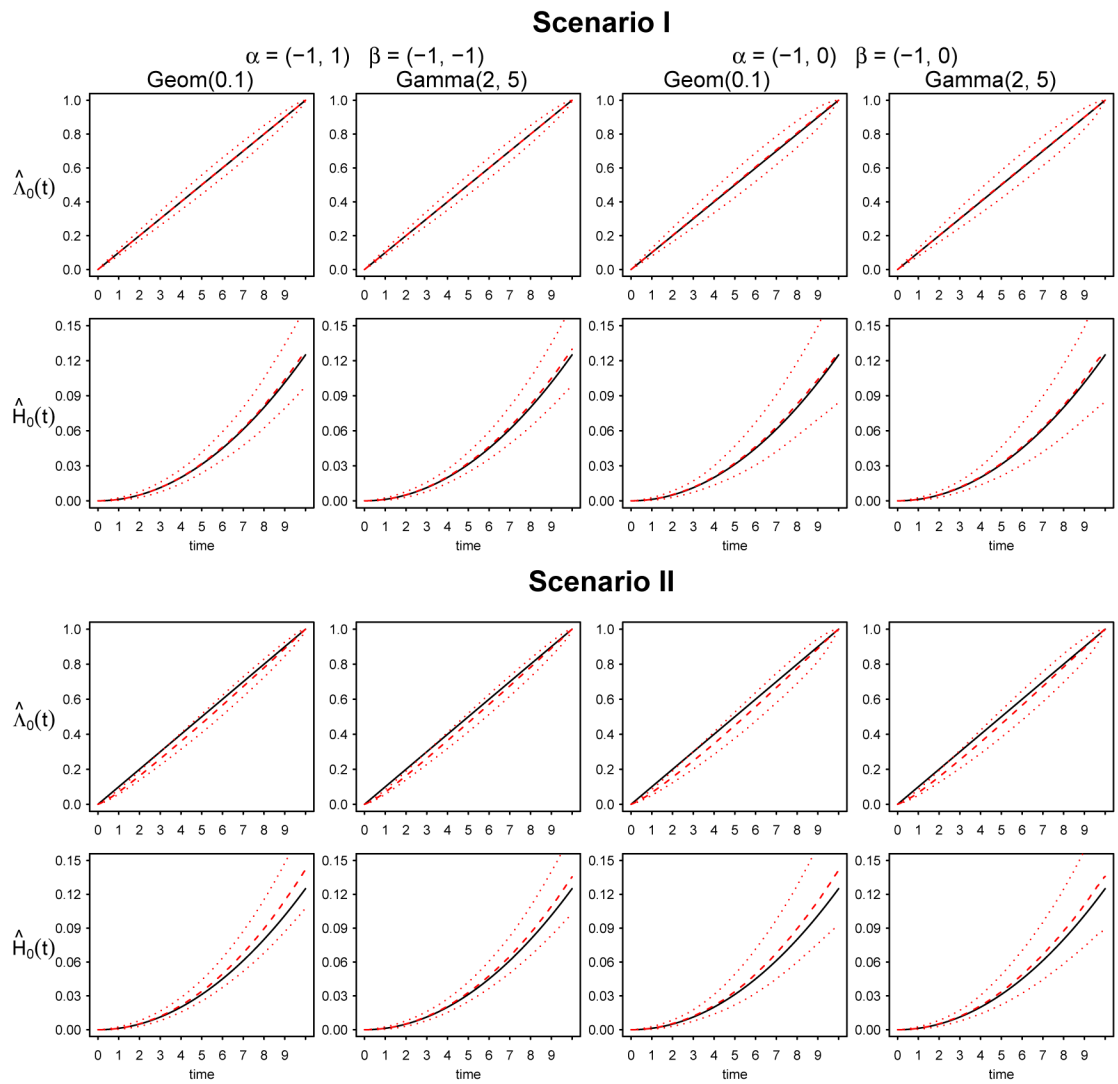


Figure 2.

Plots of estimated $\hat{\Lambda}_0(t)$ and $\hat{H}_0(t)$ with pointwise 95% confidence intervals for $n = 500$ (—, true curve; ---, empirical average; ----, pointwise 95% confidence intervals). Conditioning on Z , the recurrent process is a Poisson process and a non-Poisson process for Scenario I and Scenario II, respectively.

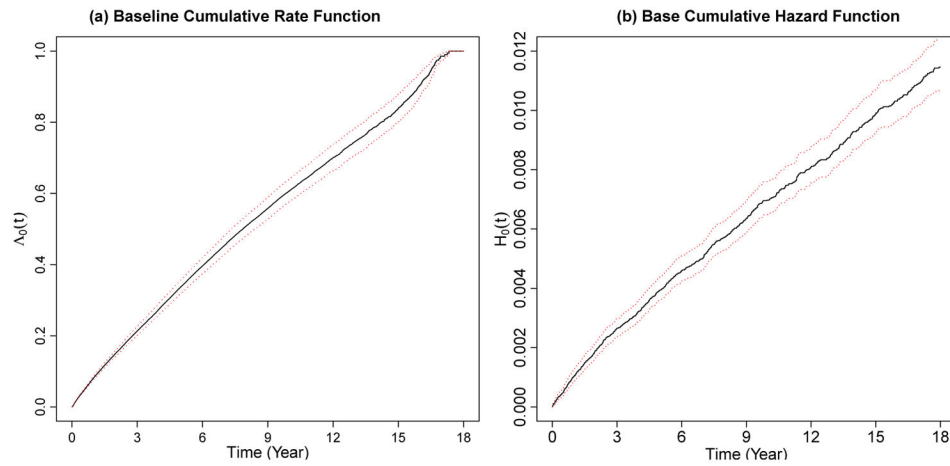


Figure 3. Plots of estimated $\hat{\Lambda}_0(t)$ and $\hat{H}_0(t)$ for the Danish Psychiatric Central Register data. (a) Baseline cumulative rate function; (b) baseline cumulative hazard function; (—, estimates; ----, pointwise 95% confidence intervals).

Table 1

Summary statistics of the simulated data with $n = 200$: Bias is the empirical bias ($\times 1000$); ESE is the empirical standard error ($\times 1000$); ASE is the average standard error ($\times 1000$); CP is the empirical coverage probability (%); Conditioning on Z , the recurrent process is a Poisson process and a non-Poisson process for Scenario I and Scenario II, respectively.

$\alpha = (-1, 1)^T, \beta = (-1, 1)^T$										$\alpha = (-1, 0)^T, \beta = (-1, 0)^T$									
$Z \sim \text{Geom}(0.1)$					$Z \sim \text{Gamma}(2, 5)$					$Z \sim \text{Geom}(0.1)$					$Z \sim \text{Gamma}(2, 5)$				
Bias	ESE	ASE	CP	CP	Bias	ESE	ASE	CP	CP	Bias	ESE	ASE	CP	CP	Bias	ESE	ASE	CP	CP
Scenario I																			
Proposed																			
a_1	23	254	250	94.7	17	211	212	95.9	42	277	271	93.0	15	259	234	93.7			
a_2	20	419	404	94.4	14	363	354	95.2	-5	460	435	94.2	-6	424	388	94.5			
β_1	6	224	246	97.0	1	213	243	98.0	4	327	313	91.2	24	377	358	91.8			
β_2	-17	371	388	97.0	-14	354	398	98.4	6	490	476	96.1	6	448	460	96.6			
Ghosh and Lin																			
a_1	911	639	596	55.6	657	683	630	74.5	793	452	453	50.8	576	540	540	75.1			
a_2	173	488	472	94.6	168	442	452	96.5	36	528	525	97.9	19	541	534	96.9			
β_1	431	264	261	58.6	310	253	244	70.7	477	333	346	66.3	323	343	335	75.5			
β_2	163	470	448	94.3	164	432	424	95.4	4	560	555	97.2	18	542	536	97.3			
Scenario II																			
Proposed																			
a_1	20	251	250	94.3	2	206	206	95.2	55	288	271	91.9	27	259	246	93.2			
a_2	24	440	402	94.5	25	353	341	94.9	-10	470	448	93.3	-2	409	390	94.5			
β_1	21	279	243	96.8	27	259	236	98.2	35	420	379	89.9	38	402	393	92.2			
β_2	-22	372	387	96.6	-6	342	380	97.9	15	478	463	94.3	-12	480	459	95.7			
Ghosh and Lin																			
a_1	845	629	583	60.7	481	604	595	82.1	846	402	448	45.0	642	515	514	69.2			
a_2	228	510	489	95.1	222	458	461	95.0	32	534	536	98.7	64	533	537	97.3			
β_1	423	267	262	60.9	289	259	245	73.9	467	361	351	67.0	349	328	334	75.9			
β_2	211	464	456	95.3	199	421	421	96.4	6	575	570	98.2	40	535	540	96.7			

Table 2

Summary statistics of the simulated data with $n = 500$: Bias is the empirical bias ($\times 1000$); ESE is the empirical standard error ($\times 1000$); ASE is the average standard error ($\times 1000$); CP is the empirical coverage probability (%); Conditioning on Z , the recurrent process is a Poisson process and a non-Poisson process for Scenario I and Scenario II, respectively.

$\alpha = (-1, 1)^T, \beta = (-1, 1)^T$										$\alpha = (-1, 0)^T, \beta = (-1, 0)^T$									
$Z \sim \text{Geom}(0.1)$					$Z \sim \text{Gamma}(2, 5)$					$Z \sim \text{Geom}(0.1)$					$Z \sim \text{Gamma}(2, 5)$				
Bias	ESE	ASE	CP	CP	Bias	ESE	ASE	CP	CP	Bias	ESE	ASE	CP	CP	Bias	ESE	ASE	CP	CP
Scenario I																			
Proposed																			
a_1	2	184	169	94.3	9	136	138	96.6	19	202	185	94.0	12	175	156	94.8			
a_2	11	282	276	95.6	8	239	234	96.5	-1	321	299	96.3	1	290	255	95.0			
β_1	-1	142	166	95.7	-3	128	160	97.7	4	224	222	93.9	7	231	224	94.5			
β_2	0	233	275	96.4	-4	221	273	96.8	-3	283	324	96.0	-5	276	304	961			
Ghosh and Lin																			
a_1	934	567	505	52.4	500	572	532	79.3	870	289	315	24.3	659	442	408	55.0			
a_2	148	332	313	94.4	245	335	319	90.1	54	323	345	97.5	125	391	344	93.1			
β_1	431	178	167	31.0	297	165	153	51.6	463	238	220	44.6	369	253	216	56.4			
β_2	162	313	286	90.7	209	279	266	89.5	31	338	351	97.6	76	387	336	95.0			
Scenario II																			
Proposed																			
a_1	-4	179	175	95.3	1	159	138	95.4	29	214	194	92.5	25	174	155	94.4			
a_2	7	298	286	95.8	11	254	234	94.5	5	347	306	93.7	6	286	254	94.0			
β_1	18	150	159	96.5	14	155	152	96.4	38	248	220	92.8	27	274	221	94.7			
β_2	8	242	260	96.5	1	237	259	96.3	-11	304	305	95.7	4	276	291	96.0			
Ghosh and Lin																			
a_1	789	589	505	59.7	387	465	475	85.3	904	275	288	18.6	697	447	389	47.9			
a_2	209	378	329	91.4	241	329	325	89.8	51	342	343	96.8	134	340	339	93.3			
β_1	396	171	167	36.1	291	162	153	50.9	473	229	220	43.1	334	252	216	59.3			
β_2	198	333	284	86.6	223	268	265	88.0	40	347	349	96.7	102	356	334	94.0			

Table 3

Summary of Denmark Psychiatric Central Register data: PE is the point estimate; SE is the standard error; biased estimator is estimates based on AFT model with independent censoring assumption.

Risk factor	Proposed				Ghosh and Lin			
	Hospitalization		Death		Hospitalization		Death	
	PE	SE	PE	SE	PE	SE	PE	SE
Onset age	0.224	0.052	-0.911	0.194	0.016	0.323	-0.897	0.168
Gender	-0.194	0.048	-0.090	0.110	0.143	0.153	0.189	0.064